Folding in Geomaterials

Pouria Behnoudfar
ORCID ID: 0000-0003-4301-3728

School of Earth and Planetary Sciences

A thesis presented for the Degree of

Doctor of Philosophy (PhD)
of
Curtin University

September 2021

Declaration

To the best of my knowledge and belief, this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

The thesis work was conducted from Feb. 2018 to Sep. 2021 under the supervision of Professor Victor M. Calo.

Pouria Behnoudfar

Date: 26/Sep./2021

Hereby, I, Pouria Behnoudfar, declare that in collaboration with Prof. Victor Calo and Dr. Quanling Deng, we published the following papers:

- Higher-order generalize-α methods for hyperbolic problems, published in the Journal of Computer Methods in Applied Mechanics and Engineering, Elsevier.

- Split generalized-α method: A linear-cost solver for multi-dimensional second-order hyperbolic systems published in the Journal of Computer Methods in Applied Mechanics and Engineering, Elsevier.

- A variationally separable splitting for the generalized-α method for parabolic equations, published in the International Journal for Numerical Methods in Engineering, Wiley.

- High-order generalized-alpha method published in the Journal of Applications in Engineering Science 4, Elsevier.

Also, we have submitted a manuscript entitled "Higher-order generalized-α methods for parabolic problems" in the Journal of Finite Elements in Analysis and Design, Elsevier. The preprint of this work is published on Arxiv. We summarize the contributions by field of activity in the following table:

	Conception and Design	Analysis	Numerical modellings	Interpretation and Discussion	Final Approval	Total % contribution
Pouria Behnoudfar	40	40	70	40	0	55
Quanling Deng	30	50	30	40	20	35
Victor Calo	30	10	0	20	80	10
Total %	100	100	100	100	100	100

The co-authors, Prof. Victor Calo and Dr. Quanling Deng, acknowledge that these represent the contribution to the above research outputs.

Quanling Deng
Signature:

Victor Calo
Signature:

Hereby, I, Pouria Behnoudfar, declare that in collaboration with Prof. Alessandro Reali, Prof. Giancarlo Sangalli, Prof. Victor Calo, and Dr. Gabriele Loli, we have submitted the paper entitled "Explicit high-order generalized-α methods for isogeometric analysis of structural dynamics" in the Journal of Computer Methods in Applied Mechanics and Engineering, Elsevier. The preprint of this work is published on Arxiv. We summarize the contributions by field of activity in the following table:

	Conception and Design	Analysis	Numerical modellings	Interpretation and Discussion	Final Approval	Total % contribution
Pouria Behnoudfar	40	50	20	35	0	40
Gabriele Loli	30	0	80	35	0	35
Alessandro Reali	10	20	0	10	35	10
Giancarlo Sangalli	15	20	0	10	35	10
Victor Calo	5	10	0	10	30	5
Total %	100	100	100	100	100	100

The co-authors, acknowledge that these represent the contribution to the above research outputs.

Gabriele Loli
Signature:

Alessandro Reali
Signature:

Giancarlo Sangalli
Signature:

Victor Calo
Signature:

Hereby, I, Pouria Behnoudfar, declare that in collaboration with Prof. Victor Calo, Prof. Maciej Paszynskia, and Dr. Marcin Los, we published the paper entitled "Fast isogeometric solvers for hyperbolic wave propagation problems" in the Journal of Computers & Mathematics with Applications, Elsevier.

Also, we have submitted a manuscript entitled "A variational splitting for high-order linear multistep methods for parabolic equations" in the International Journal for Numerical Methods in Engineering, Wiley. We summarize the contributions by field of activity in the following table:

	Conception and Design	Analysis	Numerical modellings	Interpretation and Discussion	Final Approval	Total % contribution
Pouria Behnoudfar	45	60	0	35	0	40
Marcin Los	20	0	70	20	0	30
Maciej Paszynskia	25	30	30	25	50	25
Victor Calo	10	10	0	20	50	5
Total %	100	100	100	100	100	100

The co-authors, acknowledge that these represent the contribution to the above research outputs.

Marcin Los
Signature:

Maciej Paszynskia
Signature:

Victor Calo
Signature:

Abstract

In this study, we model the localized folding in geomaterials. For this, we utilize the plate theory to describe the deformations of the middle surface of a fold and the changes of its normal vector. Plate theory provides a theoretical framework to describe a 3D phenomenon using 2D models. This leads to cheaper computations. Therefore, we consider an elastic plate embedded in a viscoelastic framework. Then, we derive a model similar to the Swift-Hohenberg equation for predicting the folding of the plate. Later, we extend the model to simulate the folding of a viscoelastic plate. Finally, we take into account the effect of the thickness of the folded layer and develop a model for thick layers. In all of the models, we include localized softening-stiffening behaviour in the supporting framework. In the next step, we study the effect of coefficients in our models and obtain non-dimensional equations. Later, we provide our detailed analyzes on the linear stability of the models to introduce stability conditions.

Later, we develop a robust finite element method for solving the resulting differential equations. We discretise the time-dependent models using time-marching techniques. In our experience, to exploit highly accurate finite element discretization techniques (e.g. isogeometric analysis (IGA)), we need comparable resolution in time. The generalized-α method provides second-order accuracy and unconditional stability for equations with first and second derivatives in time. These techniques allow the user to control the high-frequency numerical dissipation by producing an algorithm that delivers an optimal combination of high-frequency and low-frequency dissipation. The approach provides accurate approximations in low- and high-frequency regions and has been widely deployed in many engineering applications. Despite these features, the generalized-α method was limited

to second-order accuracy in time while the higher-order Runge-Kutta and other multistep schemes (e.g., Adams-Moulton, and backward differentiation formulae (BDFs)) lack explicit control over the numerical dissipation of the high frequencies. The stability region of these higher-order multistep methods shrinks as their order increases. Therefore, unconditional stability is not possible for orders higher than two. While Runge-Kutta methods have better stability regions and deliver A-stability with higher-order accuracy, they are not self-starting and require another scheme to retrieve solutions at initial time steps.

Thus, we introduce and analyze a new class of extensions of the generalized-α methods for partial differential equations (PDEs) with first- and second-order derivatives in time. These methods deliver flexible arbitrary high-order accuracy while keeping all of the appealing features of these robust and effective methods. Proposing a single parameter, our time-marching techniques has similar structures as the original second-order scheme with control on the dissipation. That is, our high-order schemes require simple modifications of the available implementations of the generalized-α method. We introduce implicit and explicit methods for hyperbolic problems and implicit ones for parabolic models.

Finally, we develop a splitting technique that reduces the algebraic solver's computational cost while maintaining the approximation's accuracy. Implicit time marching schemes for finite elements or IGA discretizations solve a matrix system at each step; this step is the solver's significant cost. We substitute the original discretization with an approximating splitting technique to significantly reduce the computation cost of the solution. We introduce a variational splitting for parabolic and hyperbolic problems using tensor-product grids to formulate the variational formulation for multi-dimensional problems. We adapt operator splitting ideas to tensor-product spatial discretizations for classical finite elements and isogeometric analysis. We write the d-dimensional formulation as a product of d 1D formulations in each dimension plus appropriate error terms. We refer to these formulations as variationally separable. Therefore, we introduce a splitting

technique to simulate the linear systems with a linear computational cost. Having sufficiently regular solutions, the approximate solution converges to the exact solution with optimal rates in space and time while reducing the computational cost significantly.

To Pantea, Sahar, Hossein, Parsa, and Parham.

Acknowledgements

First of all, I would like to gratefully thank Pantea for her endless kindness, support, and encouragement as my best friend and spouse. Also, I appreciate the continuing support of Prof. Victor Calo as supervisor and mentor. Furthermore, I thank my family for their support and love. Without them, this thesis would not have been possible.

Discussions with Dr Quanling Deng, Dr Luis Espath, Prof. Maciej Paszunski, Prof.Alessandro Reali, Prof.Giancarlo Sangalli, and Prof. Andrew Putnis have been invaluable throughout my studies.

Lastly, I am grateful for the financial support through the Research Training Program (RTP) scholarship and Tiger travel grant.

Contents

List of Figures

LIST OF FIGURES

1

Introduction

1.1 Layered-media deformation

Geological structures show various complex behaviours; the study and simulation of their evolution are key to understanding geological phenomena such as plate tectonics, rock properties, and their stress fields. Therefore, this topic attracts the attention of theoreticians, modellers, and experimentalists. For example, different settings may be considered when describing rock deformation such as the effects of microstructures [107], structural discontinuities [25, 104], fluid interactions and induced pressure [33, 40], pressure solution [82], stress corrosion [106], and different mineral reactions [6, 35, 101]. These studies consider various deformation mechanisms and describe the evolution of deformed rocks both experimentally and numerically.

A significant example of these deformation mechanisms is when geological structures are subject to forces induced by tectonic plate motion. These structures may be divided into many categories and show complex features. Folds, as an example, depict a wide range of deformations, length distributions, and patterns, including periodic folds, chevron folds, and box folds [62, 80]. Besides, localized deformations are present in most of these structures. Therefore, describing the buckling of rocks subject to geological forces is a vital and demanding task that describes the various factors that gain importance during the initiation and formation of a fold. Given the complex interactions between these different aspects, we believe that simulations can improve our understanding of the geological history of a region, material properties of the folded layers, the embedded parts and their stress fields, and the physical conditions that lead

to the formation of the observed structures.

The initial studies simulated the deformation of a layer and its embedding framework with different rheological properties. The results present folds with the periodical pattern but do not describe most of the observed folds. For example, Biot in [14] made the first attempt by analysing a single elastic rod and plate surrounded by a viscous material and concluded that the deriving parameter for buckling is the difference between the rheological properties of the embedded rod and surrounding framework. Accordingly, to determine the wavelength of the maximum amplification, he introduced the dominant wavelength λ_D as

$$\frac{\lambda_D}{h} = 2\pi \left(\frac{\mu_L}{6\mu_M} \right)^{1/3}, \tag{1.1.1}$$

with h, λ_L, λ_M, being the folded layer thickness and the viscosities of the layer and the matrix, respectively. The model (1.1.1) estimates the folded layer's viscosity from its thickness and its wavelength. Later, Sherwin & Chapple used Biot's model to analyze 800 single layer fold specimens; nevertheless, due to its simple assumptions, the model could not explain the observations [89]. Then, to improve the model, several modifications were introduced by adding nonlinear viscous materials or large amplitude folding [42, 86, 90, 91]. However, the absence of localized buckling is still an important aspect to be addressed [20, 51]. Subsequently, nonlinear folding models were proposed. For example, an alternative analyzes localized folds by studying a strut on a nonlinear elastic foundation, which buckles the strut locally if the foundation softens as it deforms [58]. Such softening behaviour may be due to the material's nonlinear nature or the multi-physics interaction [52].

Spatially localized structures and deformations are relevant to diverse fields such as reaction-diffusion systems [67], liquid crystals [61, 79], and binary fluid flows [65]. Recently, higher-order partial differential equations (PDEs) (e.g., equations with spatial derivative of equal or higher order than four) such as the Swift-Hohenberg model or other phase-field models have been employed to study localized phenomena. Some of the examples are models describing binary alloys [64], fracture propagation [63, 66], phase transition [103], evolution of microstructures [92], solidification [36], and minerals undergoing chemo-mechanical [27].

Our folding model is similar to the Swift-Hohenberg model. The general form of the Swift-Hohenberg model reads:

$$\dot{u} = ru - (q_c^2 + \Delta)^2 u + N(u; s), \tag{1.1.2}$$

where $u(x_1, x_2, t) \in \mathbb{R}$ is the studied variable with a finite wave number q_c^2 at onset, $r < 0 < s$ are parameters, and $N(u; s)$ denotes nonlinear terms. The most frequently used non-linear terms are:

$$N(u; s) = \begin{cases} f_{35} = b_3 u^3 - b_5 u^5 \\ f_{23} = c_2 u^2 - c_3 u^3 \end{cases} \tag{1.1.3}$$

Standard modelling strategies for pattern formation near instabilities use (1.1.2) with non-trivial finite spatial wavelength [54]. In the context of buckling of a layer, the fourth-order term in (1.1.2) is related to bending, while the second-order one is to the axial load. The non linearity of the foundation in (1.1.3) represents the softening and stiffening of the foundation reactions to the layer displacement.

Herein, we derive (1.1.2) using the plate theory to model the deformations of the middle surface of a plate and the changes of its normal vector. In theory of elasticity, bodies bounded by two closely-spaced surfaces with a small thickness compared to the planar dimensions are called plates (flat) or shells (curved) [81]. Such structures are common in aerospace and biomechanical (see [74]) engineering applications. Modelling thin structures with a three-dimensional continuum is expensive computationally, mainly when considering nonlinear and transient effects [13]. Therefore, modelling such structures using their 2D mid surface is often preferred in science and engineering applications. We derive the model from describing a 1D rod or a 2D plate embedded in a viscoelastic framework. Next, we modify the model to consider the plate as a viscoelastic material. Finally, we take into account the thickness of the folded layer and describe thick folds.

Later, we develop a robust finite element method for solving the resulting differential equations. We discretize the time-dependent models using time-marching techniques. We propose an entirely new class of high-order generalized-α methods. Chung and Hulbert in [26] proposed the generalized-α method for solving hyperbolic equations arising in structural dynamics and wave propagation. The method provides second-order accuracy in time and unconditional stability. The generalized-α method

extends the well-known HHT-α method of Hilber, Hughes, Taylor [50] and the WBZ-α method of Wood, Bossak, and Zienkiewicz [105]. That is, by selecting specific values for the user-defined parameters of the generalized-α method, the resulting technique reduces to either the HHT-α or WBZ-α methods. Later, the generalized-α method was extended to computational fluid dynamics, which are governed by parabolic differential equations such as the Navier-Stokes equations in [60]. These techniques allow the user to control the high-frequency numerical dissipation by producing an algorithm that delivers an optimal combination of high-frequency and low-frequency dissipation. That is, the approach provides accurate approximations in both low- and high- frequency regions; see [26, 60]. Consequently, the method has been successfully used in a wide range of engineering applications [4, 5, 46, 47, 85].

Despite these features, to date, the generalized-α method was limited to second-order accuracy in time while the higher-order Runge-Kutta and other multistep schemes (e.g., Adams-Moulton, and backward differentiation formulae (BDFs)) lack explicit control over the numerical dissipation of the high frequencies (see [2, 22, 23]). Notably, the stability region of these higher-order multistep methods shrinks as their order increases. Therefore, unconditional stability is not possible for orders higher than two. While Runge-Kutta methods have better stability regions and deliver A-stability with higher-order accuracy, they are not self-starting and require another scheme to retrieve solutions at initial time steps.

We proposed and analyzed a new class of extensions of the generalized-α methods for parabolic and hyperbolic PDEs to deliver flexible arbitrary high-order accuracy while keeping all of the appealing features of these robust and effective methods. We provide a detailed analysis of the stability behaviour of the methods and the proof for the accuracy in the temporal domain. These new methods have many applications in modelling geological phenomena such as wave propagation, reaction-diffusion, and fluid flow. Nevertheless, the applicability of these new class of higher-order methods is not limited to geological applications; effectively, these new methods can be applied to problems where the standard second-order generalized-α methods proved successful [1, 11, 16, 18, 19, 37, 85].

Next, we develop a mathematical technique that reduces the computational cost of the algebraic solver while keeping the accuracy of the approximation. Implicit time marching schemes for PDE discretizations such as finite differences, finite elements, and

isogeometric analysis solve a matrix system at each time step; this step is the solver's major computational cost. We substitute the original discretization with an approximating splitting technique [83] to significantly reduce the computation cost of the solution. For example, two standard procedures that reduce the dimension of the matrices are implicit numerical schemes [69, 83] and domain decomposition methods [84, Chapter 8]. For many PDE problem configurations, the resulting time discretization of the parabolic and hyperbolic operators can be split in a sum of simpler operators that require the solution of a simpler set of much simpler linear problems. Classic examples of such schemes are provided in the pioneering works developing direction-splitting schemes for linear parabolic problems [34, 77]. Recently, [43, 44, 70, 71] show that alternating direction splitting of tensor-product spaces delivers a linear computational cost at every time step for various PDE problems. See [73, 98] for comprehensive reviews of splitting techniques.

We propose a variational splitting for parabolic and hyperbolic problems using tensor-product meshes to derive the variational formulation for multi-dimensional problems. We use operator splitting ideas to tensor-product spatial discretizations for classical finite elements and isogeometric analysis. That is, we formulate the d-dimensional equations as a product of d 1D formulations in each dimension plus appropriate error terms. We refer to these formulations as variationally separable. Exploiting the variational separability, we present a splitting technique to solve the linear systems with a linear computational cost with respect to the degrees of freedom.

The rest of this thesis is organized as follows. Chapter 2 derives the localized folding model. We develop a model similar to the Swift-Hohenberg to study the folding of an elastic layer embedded in a viscoelastic framework. Then, we derive a folding model for a viscoelastic layer. Finally, we discuss the modelling of the folding of a thick viscoelastic layer. Also, we detail our analysis for dimensional study and linear stability of the introduced models. In Chapter 3, we propose our higher-order time integrators. Chapter 4 discusses our new splitting techniques. We provide our concluding remarks in Chapter 5 where we summarize our contributions and delineate future lines of work to pursue.

2

Folding

2.1 Folding of rods and plates as proxies for layered media

We model the folding layers as rods (1D layers) and plates (2D layers) to estimate their curvatures during the folding to achieve a more precise understanding. In continuum mechanics, rod, plate, and shell theories describe the mechanical behaviour of slender structural elements (i.e., a small thickness compared to their planar dimensions [97]). In 1766, Euler pioneered the study on the plates and was further extended by Bernoulli. Euler modelled the free vibration of rectangular and circular membranes. However, the model excluded the torsional resistance of the plate, and consequently, only a semblance and not a general agreement can be found between the results and the experiments. Kirchhoff (1824-1887) proposed the theory of the extended plate. He considered bending and stretching as well. Hence, the nonlinear terms are not neglected in his approach [94]. In general, we determine the deformation and stresses in a 2D rod or a 3D plate by tracking its centreline motion and the changes it induces on its normal vector [15, 81].

2.2 Folding of an elastic rod

We now consider an inextensible elastic layer embedded in a viscous framework. We model the effect of the structure's surrounding as a reaction from its foundation. We use a Kelvin–Voigt viscoelastic framework to model this response. For this, we utilize Winkler's idealization to represent the medium as a system of independent, closely

spaced, discrete, and linearly elastic springs and nonlinear viscous dashpots. Figure 2.1 sketches our model.

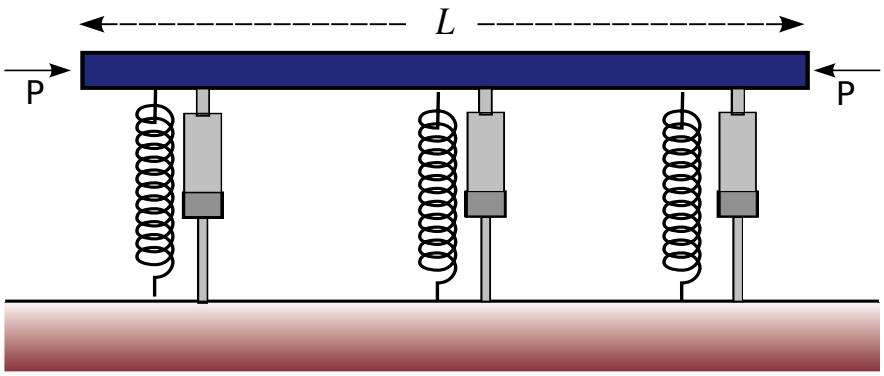


Figure 2.1: A beam surrounded by a Kelvin–Voigt viscoelastic framework.

Here, we describe the deformation of the beam due to the axial compressive force P using the vertical displacement of its centreline $w(x,t)$. Since the springs and dashpots representing the elastic and viscous behaviours, respectively, are arranged in parallel, the Kelvin–Voigt model states that the strains in each component are equal:

$$\epsilon_{total} = \epsilon_e = \epsilon_v, \tag{2.2.1}$$

where ϵ denotes the strain and subscripts e and v indicate the elastic and viscous parts, respectively. Accordingly, for the stress σ, we have:

$$\sigma_{total} = \sigma_e + \sigma_v, \tag{2.2.2}$$

which allows us to first derive the governing equations without the viscous behaviour and later add it to the model. For this, we define the beam curvature:

$$\chi = \frac{d}{dx}\arcsin w' = \frac{w''}{\sqrt{1-(w')^2}}, \tag{2.2.3}$$

where w' and w'' are the first and second differentiation of w with respect to x, respectively. Then, we derive the bending energy of a linear elastic beam as:

$$\begin{aligned} E_b &= \frac{B}{2}\int_0^L \chi^2 dx = \frac{B}{2}\int_0^L \frac{w''}{1-(w')^2} \\ &= \frac{B}{2}\int_0^L (w'')^2\left(1+(w')^2+\mathcal{O}\left((w')^4\right)\right), \end{aligned} \tag{2.2.4}$$

with B being the bending stiffness. Furthermore, the vertical deformation of the centreline leads to the shortening of the beam $\delta(w)$ which we obtain using:

$$\begin{aligned}\delta(w) &= L - \int_0^L \sqrt{1-(w')^2}dx \\ &= L - \int_0^L \left(1 - \frac{1}{2}(w')^2 + \mathcal{O}\left((w')^4\right)\right)dx,\end{aligned} \tag{2.2.5}$$

where L is the beam's length. Assuming moderate deflections, we only retain quadratic terms in equations (2.2.4) and (2.2.5). Therefore, the potential energy of the beam reads:

$$\Pi = E_b - P\delta = \int_0^L \left(\frac{B}{2}(w'')^2 - \frac{P}{2}(w')^2\right)dx. \tag{2.2.6}$$

We minimize the energy functional (2.2.6) using the Lagrangian of the system $\mathscr{L}$ (for more details, see [45]):

$$\Pi \equiv \int_0^L \mathscr{L}(w', w'')dx,$$

satisfying the Euler-Lagrange equation:

$$\frac{d^2}{dx^2}\frac{\partial \mathscr{L}}{\partial w''} - \frac{d}{dx}\frac{\partial \mathscr{L}}{\partial w'} = 0. \tag{2.2.7}$$

Substituting (2.2.6) into (2.2.7), one can readily obtain:

$$Bw'''' + Pw'' = 0. \tag{2.2.8}$$

Lastly, we require adding the framework's support as a resistive vertical force applied locally to the beam.

2.2.1 Behaviour of the foundation

In our model, we describe the elastic behaviour of the Winkler-type foundation as a localized softening and re-stiffening type using a cubic-quintic non-linear term. A time-dependent term also describes the viscous effect. Thus, we add the support of the framework as a resistive vertical force applied locally to the plate as

$$f(w) = f_e(w) + f_v(w) = k_1 w - k_2 w^3 + k_3 w^5 + \eta \dot{w}, \qquad k_i, \eta > 0 \tag{2.2.9}$$

where $f(w)$ is the vertical resistive force that depends on the local vertical deformation of the middle surface of the plate. $f_e(w)$ and $f_v(w)$ are elastic and viscoelastic effects,

respectively. $\dot{w}$ denotes the time-derivative of w. Increasing the vertical displacement w leads to softening of the foundation. This behaviour finally changes to stiffening after the deflection w reaches a specific value. We account for the foundation's elastic behaviour by rewriting (2.2.9) as a dimensionless resistive vertical force dependant on a single parameter that reads (for more details, see, section 2.3.1):

$$f(w) = w - w^3 + \mu w^5 + \dot{w}, \qquad \mu > 0. \tag{2.2.10}$$

Figure 2.2 shows the softening re-stiffening behaviour of the elastic part of the force (2.2.10) for different μ values.

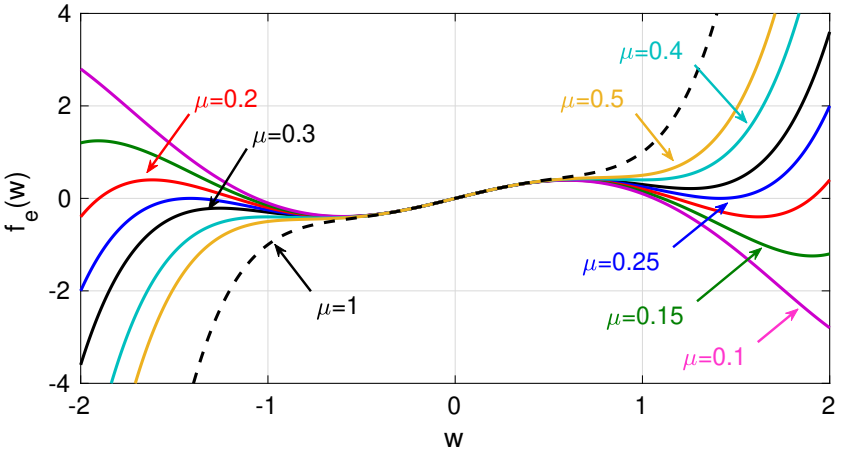


Figure 2.2: The Softening re-stiffening behaviour of elastic force f_e applied by the foundation.

In the next step, we calculate the potential energy Π_f imposed to the system from the foundation by the resistive force (2.2.9). For this, we have:

$$\Pi_f = \Pi_{f_e}(w) + \Pi_{f_v}(w) = -\int_{\Omega_0} \frac{k_1}{2}w^2 - \frac{k_2}{4}w^4 + \frac{k_3}{6}w^6 + \eta w\dot{w}, \qquad k_i, \eta > 0. \tag{2.2.11}$$

Similarly, the potential energy corresponding to the dimensionless force (2.2.10) reads:

$$\Pi_f = -\int_{\Omega_0} \frac{1}{2}w^2 - \frac{1}{4}w^4 + \frac{\mu}{6}w^6 + w\dot{w}, \qquad \mu > 0. \tag{2.2.12}$$

In figure 2.3, we show the evolution of the elastic part Π_{f_e} of the energy functional imposed by the foundation.

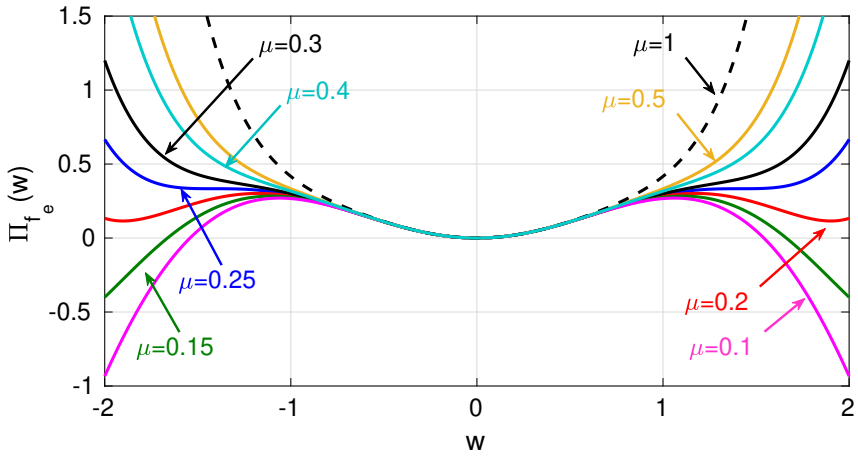


Figure 2.3: The evolution of the elastic part of the energy functional imposed by the foundation.

Remark 1. *In our model, the foundation's elastic behaviour softens and re-stiffens due to the cubic–quintic nonlinear interactions. Increasing the vertical displacement softens the foundation due to the cubic term that is energetically favourable to buckle the beam further. As this deflection progresses, the quintic term increases and eventually dominates the interaction stiffening the foundation's response. This interaction induces a trade-off between growing a fold at the current and establishing a new fold close by in a softer region. Hence, the softening and re-stiffening of the foundation induce snaking where a fold localizes and expands sequentially due to this ongoing competition of softening and stiffening.*

As a consequence, our aggregate model reads:

$$Bw'''' + Pw'' + k_1 w - k_2 w^3 + k_3 w^5 + \eta \dot{w} = 0. \tag{2.2.13}$$

Next, we extend this model to predict the folding of a 2D plate.

2.3 Folding of a plate

Herein, we extend the derived model in the previous section to a 2D model. For this, we start by defining the domain Ω:

$$\Omega := \left\{ (x_1,\, x_2,\, x_3) \in \mathbb{R}^3 : x_1,\, x_2 \in \Omega_0,\, x_3 \in [-h/2\,,\, h/2] \right\},$$

with $\Omega_0 := (0, L_1) \times (0, L_2)$ being the middle surface of the plate and h denoting the plate thickness. We consider compressive forces P_1 and P_2 along x_1 and x_2, respectively,

are applied on the edges of Ω. Similarly, the plate is supported using a viscoelastic foundation. we present our model in figure 2.4. Then, assuming $h << \min(L_1, L_2)$,

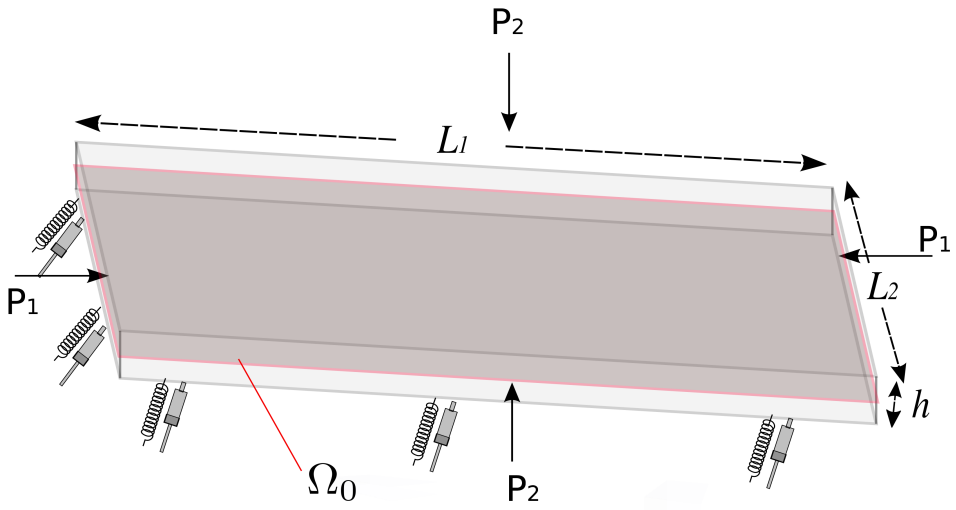


Figure 2.4: A 2D plate embedded in a viscoelastic framework.

the displacement field reads:

$$\mathbf{u}(x_1, x_2, x_3) = \left(-x_3 \frac{\partial w(x_1, x_2)}{\partial x_1}, -x_3 \frac{\partial w(x_1, x_2)}{\partial x_2}, w(x_1, x_2) \right), \tag{2.3.1}$$

where $w(x_1, x_2)$ is the deflection along x_3. Using the von–Karman strain-–displacement relations, the non-zero strain components at an arbitrary point of the plate are related to the displacement field of the middle surface (2.3.1), the changes in the curvatures k_1 and k_2, and torsion k_{12} of the middle surface as [24]:

$$\epsilon_1 = x_3\, k_1, \qquad \epsilon_2 = x_3\, k_2, \qquad \gamma_{12} = x_3\, k_{12}. \tag{2.3.2}$$

Following closely the Kirchhoff–Love theory of plates, the rotation of the normal to the mid-surface reads [72]:

$$\zeta_1 = \frac{\partial w(x_1, x_2)}{\partial x_1}, \qquad \zeta_2 = \frac{\partial w(x_1, x_2)}{\partial x_2}, \tag{2.3.3}$$

where ζ_1 and ζ_2 are the rotation with respect to the directions x_1 and x_2, respectively. Figure 2.5 sketches the details. Thus, one can readily obtain the strain components as:

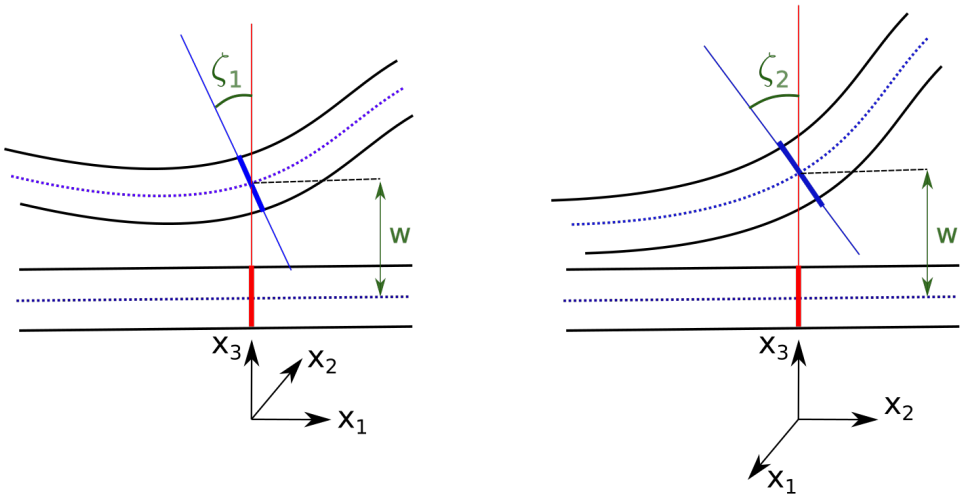


Figure 2.5: Normal vector rotation of the middle surface. Blue dotted line is the position of the middle surface.

$$\begin{aligned}
\epsilon_1 &= -x_3 \frac{\partial^2 w(x_1, x_2)}{\partial x_1^2}, \\
\epsilon_2 &= -x_3 \frac{\partial^2 w(x_1, x_2)}{\partial x_2^2}, \\
\gamma_{12} &= -2x_3 \frac{\partial^2 w(x_1, x_2)}{\partial x_1 \partial x_2}.
\end{aligned} \tag{2.3.4}$$

Remark 2. *The assumption of* $h << \min(L_1, L_2)$ *and consequently the displacement field* (2.3.1) *lead to neglect of the transverse shear strains:*

$$\gamma_{13} = \gamma_{23} = 0. \tag{2.3.5}$$

Following the elastic theory, the strain energy of the body is then proportional to

the functional:

$$\Pi_S(\boldsymbol{u}) = \int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{u}) = \int_{\Omega_0} \int_{-h/2}^{h/2} \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{u}) = \int_{\Omega_0} \int_{-h/2}^{h/2} (\sigma_1\epsilon_1 + \sigma_2\epsilon_2 + \sigma_{12}\epsilon_{12}) \tag{2.3.6}$$

where $\boldsymbol{\sigma}$ and $\boldsymbol{\epsilon}$ are stress and strain tensors, respectively. Later, we use a Linear constitutive relation, such that:

$$\boldsymbol{\sigma} = \mathbb{E}\,\boldsymbol{\epsilon}, \tag{2.3.7}$$

which can be written element-wise as:

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{23} \end{bmatrix} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 & 0 & 0 \\ \nu & 1 & 0 & 0 & 0 \\ 0 & 0 & \dfrac{1-\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & \dfrac{1-\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & \dfrac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \gamma_{12} \\ \gamma_{13} \\ \gamma_{23} \end{bmatrix}. \tag{2.3.8}$$

Thus, we derive the strain energy due to plate bending and twisting as:

$$\Pi_S(w) = \int_{\Omega_0} \left\{ -M_1 \frac{\partial^2 w}{\partial x_1^2} - M_2 \frac{\partial^2 w}{\partial x_2^2} - 2M_{12} \frac{\partial^2 w}{\partial x_1 \partial x_2} \right\}, \tag{2.3.9}$$

where M terms are the horizontal stress resultants defined by:

$$M_1 = \int_{-h/2}^{h/2} z\sigma_1\, dz, \qquad M_2 = \int_{-h/2}^{h/2} z\sigma_2\, dz, \qquad M_{12} = \int_{-h/2}^{h/2} z\sigma_{12}\, dz. \tag{2.3.10}$$

Finally, we have:

$$\begin{aligned} \Pi_S(w) &= \int_{\Omega_0} \frac{D}{2} \left[\left(\frac{\partial^2 w}{\partial x_1^2} \right)^2 + 2\nu \frac{\partial^2 w}{\partial x_1^2} \frac{\partial^2 w}{\partial x_2^2} + 2(1-\nu) \left(\frac{\partial^2 w}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 w}{\partial x_2^2} \right)^2 \right] \\ &= \int_{\Omega_0} \frac{D}{2} \left[\left(\frac{\partial^2 w}{\partial x_1^2} + \frac{\partial^2 w}{\partial x_2^2} \right)^2 - 2(1-\nu) \left(\frac{\partial^2 w}{\partial x_1^2} \frac{\partial^2 w}{\partial x_2^2} - \left(\frac{\partial^2 w}{\partial x_1 \partial x_2} \right)^2 \right) \right], \end{aligned} \tag{2.3.11}$$

where D is

$$D := \int_{-h/2}^{h/2} z^2 \frac{E}{1-\nu^2} dz = \frac{Eh^3}{12(1-\nu^2)}. \tag{2.3.12}$$

with ν being the Poisson's ratio. Next, we require to determine the work done by the applied axial forces to the unit of thickness $\tilde{P}_i = P_i/h$, $i = 1, 2$. For this, we calculate

the shifting of the edges of the plate horizontally which leads to the shortenings δ_i as:

$$\begin{aligned}\delta_1(w) &= L_1 \times h - \int_{\Omega_0}\int_{-h/2}^{h/2}\sqrt{1-\left(\frac{\partial w}{\partial x_1}\right)^2},\\ \delta_2(w) &= L_2 \times h - \int_{\Omega_0}\int_{-h/2}^{h/2}\sqrt{1-\left(\frac{\partial w}{\partial x_2}\right)^2}.\end{aligned} \tag{2.3.13}$$

Using a Taylor's expansion and the approximation $\sqrt{1-\alpha} \approx 1 - \frac{\alpha}{2}$ for $\alpha << 1$, we approximate (2.3.13) as

$$\begin{aligned}\delta_1(w) &= L_1 \times h - \int_{\Omega_0}\int_{-h/2}^{h/2} 1 - \frac{1}{2}\left(\frac{\partial w}{\partial x_1}\right)^2 + \mathcal{O}\left(\left(\frac{\partial w}{\partial x_1}\right)^4\right),\\ \delta_2(w) &= L_2 \times h - \int_{\Omega_0}\int_{-h/2}^{h/2} 1 - \frac{1}{2}\left(\frac{\partial w}{\partial x_2}\right)^2 + \mathcal{O}\left(\left(\frac{\partial w}{\partial x_2}\right)^4\right).\end{aligned} \tag{2.3.14}$$

Thus, the work done by the forces can be written as:

$$\Pi_p(w) = -h\int_{\Omega_0}\frac{\tilde{P}_1}{2}\left(\frac{\partial w}{\partial x_1}\right)^2 + \frac{\tilde{P}_2}{2}\left(\frac{\partial w}{\partial x_2}\right)^2. \tag{2.3.15}$$

One can readily rewrite (2.3.15) as

$$\Pi_p(w) = -\int_{\Omega_0}\frac{P_1}{2}\left(\frac{\partial w}{\partial x_1}\right)^2 + \frac{P_2}{2}\left(\frac{\partial w}{\partial x_2}\right)^2 \tag{2.3.16}$$

Therefore, from (2.3.11) and (2.3.15), we obtain the potential energy of the plate as:

$$\Pi(w) = \Pi_S(w) + \Pi_p(w) = \int_{\Omega_0}\mathscr{L}\left(x_1, x_2, \nabla(w), \boldsymbol{H}(w)\right), \tag{2.3.17}$$

with ∇ denoting the spatial gradient, e.g. $\nabla(u) = \left[\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}\right]^T$. $\boldsymbol{H}$ is the Hessian matrix with entries of $(\boldsymbol{H}_u)_{i,j} = \dfrac{\partial^2 u}{\partial x_i \partial x_j}$. Following a similar argument, to minimize the functional energy (2.3.17), we use an Euler-Lagrange equation that reads [102]:

$$\begin{aligned}\frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial w_{,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial w_{,11}}\right)\\ - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial w_{,22}}\right) = 0.\end{aligned} \tag{2.3.18}$$

where

$$w_{,i} := \frac{\partial w}{\partial x_i}, \qquad w_{,ij} := \frac{\partial^2 w}{\partial x_i \partial x_j}, \qquad i,j = 1,2. \tag{2.3.19}$$

Substituting (2.3.17) into (2.3.18), it is straightforward to obtain:

$$D\left[\frac{\partial^4 w}{\partial x_1^4}+\frac{\partial^4 w}{\partial x_2^4}+2\frac{\partial^4 w}{\partial x_1^2\partial x_2^2}\right]+P_1\frac{\partial^2 w}{\partial x_1^2}+P_2\frac{\partial^2 w}{\partial x_2^2}=0. \tag{2.3.20}$$

Finally, adding the viscoelastic response of the foundation and setting $P = P_1 = P_2$, we derive the model as:

$$D\Delta^2 w + P\Delta w + k_1 w - k_2 w^3 + k_3 w^5 + \eta\dot{w} = 0, \tag{2.3.21}$$

where Δ is the Laplacian operator defined as $\Delta w = \dfrac{\partial^2 w}{\partial x_1^2} + \dfrac{\partial^2 w}{\partial x_2^2}$. In the next section, we consider (2.3.21) for further analysis.

2.3.1 Generalized homogeneity π theorem

Here, we aim to derive the dimensionless version of the model we derived for the beam folding (2.2.13) and a plate (2.3.21). For this aim, we use Buckingham π theorem for transformation to dimensionless parameters [30]. This theorem reduces a problem in determining [49]:

$$a = \mathscr{F}\left(a_1, \cdots, a_k, b_1, \cdots, b_m\right), \tag{2.3.22}$$

where a is a quantity being determined with $k+m$ arguments and a_i denotes arguments with independent dimensions. b_i represents arguments with dimensions being expressed using dimensions of a_i as:

$$[b_i] = [a_1]^{p_i} \cdots [a_k]^{r_i} \tag{2.3.23}$$

Then, introducing the variables

$$\begin{aligned}
\pi &= \frac{a}{a_1^p \cdots a_k^r},\\
\pi_i &= \frac{b_i}{a_1^{p_i} \cdots a_k^{r_i}},\\
\pi_m &= \frac{b_m}{a_1^{p_m} \cdots a_k^{r_m}},
\end{aligned}$$

allows us to formulate:

$$\pi = \frac{\mathscr{F}\left(a_1, \cdots, a_k, b_1, \cdots, b_m\right)}{a_1^p \cdots a_k^r}. \tag{2.3.24}$$

Thus, from (2.3.22) and (2.3.24), we have:

$$\pi = \mathcal{F}\left(a_1, \cdots, a_k, \pi_1, \cdots, \pi_m\right), \tag{2.3.25}$$

with $\mathcal{F}$ being a certain function.

Next, we show that changing a_1 by an arbitrary factor does not lead to any changes in $a_2, \cdots, a_k$ and $\pi_1, \cdots, \pi_m$. Similarly, this is valid for changing a_i by an arbitrary factor. Therefore, $\mathcal{F}$ is independent of $a_1, \cdots, a_k$, and we can reformulate (2.3.24) as:

$$\pi = \Phi\left(\pi_1, \cdots, \pi_m\right). \tag{2.3.26}$$

From (2.3.24) and (2.3.26), one can readily obtain:

$$\mathscr{F}\left(a_1, \cdots, a_k, b_1, \cdots, b_m\right) = a_1^p \cdots a_k^r \Phi\left(\frac{b_1}{a_1^{p_1} \cdots a_k^{r_1}}, \cdots, \frac{b_m}{a_1^{p_m} \cdots a_k^{r_m}}\right). \tag{2.3.27}$$

We use a similar logic to (2.3.27) in the rest of the discussion to derive the non-dimensional models for beam and plate folding.

We derive the dimensionless version of the model (2.3.21), by closely following (2.3.22) and (2.3.24) and defining the independent arguments $a_1 = w$, $a_2 = t$, $a_3 = x_1$, $a_4 = x_2$, $a_5 = P$, and $a_6 = \mu$. This allows us to rewrite (2.3.21) as:

$$\Delta^2 w + P\Delta w + w - w^3 + \mu w^5 + \dot{w} = 0. \tag{2.3.28}$$

by setting:

$$\begin{aligned} x_1 &\to \left(\frac{D}{k_1}\right)^{1/4} x_1, & x_2 &\to \left(\frac{D}{k_1}\right)^{1/4} x_2, \\ w &\to \sqrt{\frac{k_1}{k_2}} w, & t &\to \frac{\eta}{k_1} t \\ P &\to \frac{P}{\sqrt{D k_1}}, & \mu &\to \frac{k_1 k_3}{k_2^2} \end{aligned}$$

Later, we study the stability behaviour of the model (2.3.28).

2.3.2 Linear-stability analysis

We analyse the behaviour of the models (2.3.28) by rewriting them as:

$$\dot{w} := \Psi(w). \tag{2.3.29}$$

We first find the fixed points of the models in (2.3.28). For this, we require to find the stationary point of the energy functional, which is equivalent to the zeros of the resistive force $f_e(w)$. Thus, we have:

$$\begin{aligned} w_0 &= 0, \\ w_{\pm 1} &= \pm\sqrt{\frac{1-\sqrt{1-4\mu}}{2\mu}}, \\ w_{\pm 2} &= \pm\sqrt{\frac{1+\sqrt{1-4\mu}}{2\mu}}. \end{aligned} \tag{2.3.30}$$

Remark 3. *From* (2.3.30)*, we conclude that to have softening and re-stiffening behaviour in the framework, one requires to choose* $0 < \mu \leq 0.25$*, as Figure 2.2 shows.*

Then, linearizing the non-linear model in the vicinity of the fixed points $w_\star$ using small perturbations, we obtain:

$$w(\cdot\,,\,t) = w_\star + \delta w(\cdot\,,\,t). \tag{2.3.31}$$

Inserting the perturbation ansatz (2.3.31) into (2.3.29), we have

$$g\left(w_\star + \delta w(\cdot\,,\,t)\right) \approx g(w_\star) + g'(w_\star)\delta w(\cdot\,,\,t) = g'(w_\star)\delta w(\cdot\,,\,t) \tag{2.3.32}$$

Thus, we find that a linear ODE governs the growth of the perturbation $\delta w(\cdot\,,\,t)$, that is

$$\frac{d}{dt}\delta w(\cdot\,,\,t) = g'(w_\star)\delta w(\cdot\,,\,t) \tag{2.3.33}$$

with the solution of

$$\delta w(\cdot\,,\,t) = \delta w(\cdot\,,\,0)\exp\left(g'(w_\star)t\right). \tag{2.3.34}$$

Therefore, if $g'(w_\star) > 0$, the perturbation grows in the temporal domain, and the fixed point is linearly unstable. Whereas, for $g'(w_\star) > 0$, the fixed point shows stable behaviour, and the perturbation decays with the limit value of

$$\lim_{t\to\infty} \delta w(\cdot\,,\,t) = 0. \tag{2.3.35}$$

For further analysis, we consider a perturbation function:

$$\delta w(\cdot\,,\,t) = \epsilon\exp\left(\sigma t - i\mathbf{k}\cdot\mathbf{x}\right), \tag{2.3.36}$$

where $\mathbf{x} := (x_1, x_2)$ and $\mathbf{k} := (k_1, k_2)$ is defined as the vector of modes in x_1 and x_2 directions. Plugging (2.3.36) into the linearized form of the model, the dispersion relations for the model (2.3.28) read:

$$\begin{aligned}
\sigma_0(\mathbf{k}) &= -\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 + 1\right), \\
\sigma_{\pm 1}(\mathbf{k}) &= -\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 - 4\right) \mp \frac{1}{\mu}\left(1 - \sqrt{1-4\mu}\right), \\
\sigma_{\pm 2}(\mathbf{k}) &= -\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 - 4\right) \mp \frac{1}{\mu}\left(1 + \sqrt{1-4\mu}\right),
\end{aligned} \tag{2.3.37}$$

with $|\mathbf{k}| = \sqrt{k_x^2 + k_y^2}$ being the magnitude of the vector $\mathbf{k}$.

As we discussed before, the k-modes with $\sigma > 0$ lead to instability. Therefore, the dispersion relations in (2.3.37) show that the model have stable behaviour around the point ω_0 when the applied force P is compressional. Around the other fixed points $w_{\pm 1}$ and $w_{\pm 2}$, the solution is stable for sufficiently large modes regardless of μ and P. To analyse the stability behaviour for general modes, we consider the lowest mode $|\mathbf{k}| = 1$. Then, one requires to have ξ_1, ξ_2, ξ_3, $\xi_4 \geq 0$ for compressive P to have stability around the fixed points σ_{+1}, σ_{-1}, σ_{+2}, and σ_{-2}, respectively. Here, we define:

$$\begin{aligned}
\xi_1 &= -4\mu + 1 - \sqrt{1-4\mu}, \\
\xi_2 &= -4\mu - 1 + \sqrt{1-4\mu}, \\
\xi_3 &= -4\mu + 1 + \sqrt{1-4\mu}, \\
\xi_4 &= -4\mu - 1 - \sqrt{1-4\mu}.
\end{aligned} \tag{2.3.38}$$

Figure 2.6.a presents the variations of ξ_j, $j = 1, \cdots, 4$ with respect to μ. We can see that only the fixed point σ_{+2} is stable with any compressive axial force P. For other cases, we derive P_{min} from (2.3.37) and require to set $|P| > P_{min}$ where the negative sign represents the compressibility (see, figure 2.6.b). Otherwise, the behaviour is non-trivial and depends on the values of $|k|$ and μ.

2.4 Viscoelastic plate

In this section, we extend our model further and consider that both the plate and the surrounding framework behave as a viscoelastic material. Similarly, we use the Kelvin-Voigt viscoelastic model for the plate to write the stress tensor as:

$$\boldsymbol{\sigma}(t) = \mathbb{E}\,\boldsymbol{\epsilon}(t) + \tau_c \mathbb{E} \frac{\partial \boldsymbol{\epsilon}(t)}{\partial t}, \tag{2.4.1}$$

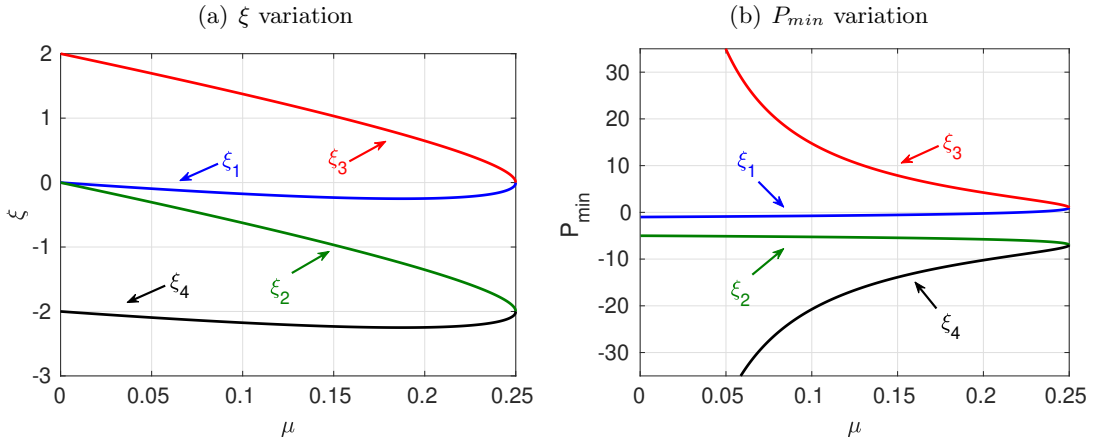


Figure 2.6: The effect of the axial force and μ on the stability of the model.

with τ_c denoting the viscoelastic relaxation parameter. Therefore, we have:

$$\sigma_1(t) = \frac{E}{1-\nu^2}\left(\epsilon_1 + \nu\epsilon_2\right) + \tau_c \frac{E}{1-\nu^2}\left(\frac{\partial \epsilon_1}{\partial t} + \nu\frac{\partial \epsilon_2}{\partial t}\right), \tag{2.4.2}$$

$$\sigma_2(t) = \frac{E}{1-\nu^2}\left(\epsilon_2 + \nu\epsilon_1\right) + \tau_c \frac{E}{1-\nu^2}\left(\frac{\partial \epsilon_2}{\partial t} + \nu\frac{\partial \epsilon_1}{\partial t}\right), \tag{2.4.3}$$

$$\sigma_{12}(t) = \frac{E}{2(1+\nu)}\epsilon_{12} + \tau_c \frac{E}{2(1+\nu)}\frac{\partial \epsilon_{12}}{\partial t}. \tag{2.4.4}$$

Thus, we can propose the energy of the plate as the summation of the elastic and viscous contributions as:

$$\Pi_S(w) = \Pi_E(w) + \Pi_V(w), \tag{2.4.5}$$

where $\Pi_E(w)$ is the energy of the elastic plate and $\Pi_V(w)$ is the viscous contribution. Following closely the previous section, from (2.3.9) and (2.4.1), we have:

$$\begin{aligned}\Pi_V(w) = \int_{\Omega_0} \tau_c D \Bigg[& \frac{\partial^2 w}{\partial x_1^2}\left(\frac{\partial^2 \dot{w}}{\partial x_1^2} + \nu\frac{\partial^2 \dot{w}}{\partial x_2^2}\right) \\ & + \frac{\partial^2 w}{\partial x_2^2}\left(\frac{\partial^2 \dot{w}}{\partial x_2^2} + \nu\frac{\partial^2 \dot{w}}{\partial x_1^2}\right) + 2(1-\nu)\left(\frac{\partial^2 w}{\partial x_1 \partial x_2}\right)\left(\frac{\partial^2 \dot{w}}{\partial x_1 \partial x_2}\right)\Bigg].\end{aligned} \tag{2.4.6}$$

Also, we obtain $\Pi_E(w)$ similar to (2.3.11). The potential energy of the plate affixed to a foundation with localized behaviour is the summation of the energies of the elastic and viscous contributions of the plate, work done by the applied axial forces, and the work of the foundation response. Thus, we define the system's potential energy as:

$$\Pi(w) = \Pi_E(w) + \Pi_V(w) + \Pi_p(w) + \Pi_f(w). \tag{2.4.7}$$

Plugging (2.2.11), (2.3.11), (2.3.16), and (2.4.6) into (2.4.7), we obtain the potential-energy functional of the system as:

$$\begin{aligned}\Pi(w) = &\int_{\Omega_0} \frac{D}{2}\left[\left(\frac{\partial^2 w}{\partial x_1^2}+\frac{\partial^2 w}{\partial x_2^2}\right)^2 - 2(1-\nu)\left(\frac{\partial^2 w}{\partial x_1^2}\frac{\partial^2 w}{\partial x_2^2} - \left(\frac{\partial^2 w}{\partial x_1 \partial x_2}\right)^2\right)\right] \\ &+ \int_{\Omega_0} \tau_c D\left[\frac{\partial^2 w}{\partial x_1^2}\left(\frac{\partial^2 \dot{w}}{\partial x_1^2}+\nu\frac{\partial^2 \dot{w}}{\partial x_2^2}\right) + \frac{\partial^2 w}{\partial x_2^2}\left(\frac{\partial^2 \dot{w}}{\partial x_2^2}+\nu\frac{\partial^2 \dot{w}}{\partial x_1^2}\right)\right. \\ &\left. + 2(1-\nu)\left(\frac{\partial^2 w}{\partial x_1 \partial x_2}\right)\left(\frac{\partial^2 \dot{w}}{\partial x_1 \partial x_2}\right)\right] - \int_{\Omega_0}\frac{P}{2}\left(\left(\frac{\partial w}{\partial x_1}\right)^2+\left(\frac{\partial w}{\partial x_2}\right)^2\right) \\ &- \int_{\Omega_0}\frac{k_1}{2}w^2 - \frac{k_2}{4}w^4 + \frac{k_3}{6}w^6 + \eta w\dot{w} = \int_{\Omega_0}\mathscr{L}\left(x_1, x_2, w, \nabla(w), \nabla(\dot{w}), \boldsymbol{H}(w), \boldsymbol{H}(\dot{w})\right).\end{aligned} \tag{2.4.8}$$

Thus, we minimize $\Pi(w)$ in (2.4.8) using an Euler-Lagrange equation that reads [102]:

$$\begin{aligned}&\frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial w_{,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial w_{,11}}\right) \\ &\quad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial w_{,22}}\right) \\ &\quad + \frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{,11}}\right) \\ &\quad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{,22}}\right) - \frac{\partial \mathscr{L}}{\partial w} = 0.\end{aligned} \tag{2.4.9}$$

Finally, we obtain the following model:

$$D\Delta^2 w + 2\tau_c D\Delta^2 \dot{w} + P_1\frac{\partial^2 w}{\partial x_1^2} + P_2\frac{\partial^2 w}{\partial x_2^2} + k_1 w - k_2 w^3 + k_3 w^5 + \eta\dot{w} = 0. \tag{2.4.10}$$

Setting $P_1 = P_2$, we have:

$$D\Delta^2 w + 2\tau_c D\Delta^2 \dot{w} + P\Delta w + k_1 w - k_2 w^3 + k_3 w^5 + \eta\dot{w} = 0. \tag{2.4.11}$$

In the next section, we conduct the dimensional analysis and derive the dimensionless version of the proposed model (2.4.11). We also bound the corresponding parameters to provide linear stability.

2.4.1 Dimensional analysis and study of linear stability

Herein, we continue the analysis of Section 2.3.1, which allows us to write the dimensionless version of (2.4.11) as:

$$\Delta^2 w + \lambda\Delta^2 \dot{w} + P\Delta w + w - w^3 + \mu w^5 + \dot{w} = 0. \tag{2.4.12}$$

Similarly, we define:

$$
\begin{aligned}
x_1 &\to \left(\frac{D}{k_1}\right)^{1/4} x_1, & x_2 &\to \left(\frac{D}{k_1}\right)^{1/4} x_2,\\
w &\to \sqrt{\frac{k_1}{k_2}} w, & t &\to \frac{\eta}{k_1} t\\
P &\to \frac{P}{\sqrt{D k_1}}, & \mu &\to \frac{k_1 k_3}{k_2^2}\\
\lambda &\to \frac{2\tau_c k_1}{\eta t}
\end{aligned}
$$

Next, we use the non-dimensional model (2.4.12) for the stability analysis. First, we find the zeros of the forcing term (see, (2.3.30)). Then, linearizing the non-linear model near the fixed points and inserting the small perturbation (2.3.31), the dispersion relations for the model (2.4.12) are:

$$
\begin{aligned}
\sigma_{0,V}(\mathbf{k}) &= \frac{-1}{1+\lambda|\mathbf{k}|^4}\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 + 1\right),\\
\sigma_{\pm_1,V}(\mathbf{k}) &= \frac{-1}{\mu\left(1+\lambda|\mathbf{k}|^4\right)}\left(\mu\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 - 4\right) \pm \left(1 - \sqrt{1-4\mu}\right)\right),\\
\sigma_{\pm_2,V}(\mathbf{k}) &= \frac{-1}{\mu\left(1+\lambda|\mathbf{k}|^4\right)}\left(\mu\left(|\mathbf{k}|^4 - P|\mathbf{k}|^2 - 4\right) \pm \left(1 + \sqrt{1-4\mu}\right)\right).
\end{aligned}
\tag{2.4.13}
$$

The dispersion relations (2.4.13) shows that the models behave similar to the model for the elastic layer around the point ω_0 and is stable when the applied force P is compressive and $\lambda \geq 0$. Around other fixed points $w_{\pm 1}$ and $w_{\pm 2}$, it is required to have $\lambda \geq 0$ and large modes. For general modes, choosing $\lambda \geq 0$, we also need ξ_1, ξ_2, ξ_3, $\xi_4 \geq 0$ for compressive P to have stable behaviour around the fixed points σ_{+1}, σ_{-1}, σ_{+2}, and σ_{-2}, respectively. Here, we refer the reader to figures (2.6).

2.4.2 Folding of thick layers

The central assumption in deriving the models in the previous sections is that the layer is considerably thin with respect to its lateral dimensions. We overcome this shortcoming for thick layers by using higher-order shear deformation theories and redefine the displacement field $\mathbf{u} = (u_1, u_2, u_3)$ as:

$$
\begin{aligned}
u_1(x_1, x_2, x_3) &= -x_3 \frac{\partial w_b(x_1, x_2)}{\partial x_1} - \frac{4x_3^3}{3h^2}\frac{\partial w_s(x_1, x_2)}{\partial x_1},\\
u_2(x_1, x_2, x_3) &= -x_3 \frac{\partial w_b(x_1, x_2)}{\partial x_2} - \frac{4x_3^3}{3h^2}\frac{\partial w_s(x_1, x_2)}{\partial x_2},\\
u_3(x_1, x_2, x_3) &= w_b(x_1, x_2) + w_s(x_1, x_2),
\end{aligned}
\tag{2.4.14}
$$

with w_b and w_s being the vertical deformations due to the bending and shear components, respectively. This higher-order theory allows us to account for shear deformation effects. Later, the strains obtained in (2.3.4) become:

$$\epsilon_1 = -x_3 \frac{\partial^2 w_b(x_1, x_2)}{\partial x_1^2} - \frac{4x_3^3}{3h^2} \frac{\partial^2 w_s(x_1, x_2)}{\partial x_1^2}, \tag{2.4.15}$$

$$\epsilon_2 = -x_3 \frac{\partial^2 w_b(x_1, x_2)}{\partial x_2^2} - \frac{4x_3^3}{3h^2} \frac{\partial^2 w_s(x_1, x_2)}{\partial x_2^2}, \tag{2.4.16}$$

$$\gamma_{12} = -2x_3 \frac{\partial^2 w_b(x_1, x_2)}{\partial x_1 \partial x_2} - \frac{8x_3^3}{3h^2} \frac{\partial^2 w_s(x_1, x_2)}{\partial x_1 \partial x_2}. \tag{2.4.17}$$

We now take into account the transverse shear strains as:

$$\gamma_{13} = \left(1 - \frac{4x_3^2}{h^2}\right) \frac{\partial w_s(x_1, x_2)}{\partial x_1}, \tag{2.4.18}$$

$$\gamma_{23} = \left(1 - \frac{4x_3^2}{h^2}\right) \frac{\partial w_s(x_1, x_2)}{\partial x_2}. \tag{2.4.19}$$

We can show that the transverse shear stresses in (2.4.18) satisfy the condition of zero transverse shear stresses at the top ($z = h/2$) and bottom ($z = -h/2$) surfaces of the plate. From (2.4.2) and (2.4.15), we calculate the general form of the strain energy as:

$$\begin{aligned} \Pi_S(w_b\,,\, w_s) &= \int_{\Omega_0} \int_{-h/2}^{h/2} (\sigma_1\epsilon_1 + \sigma_2\epsilon_2 + \sigma_{12}\epsilon_{12} + \sigma_{13}\epsilon_{13} + \sigma_{23}\epsilon_{23}) \\ &= \int_{\Omega_0} \Bigg\{ - M_1^b \frac{\partial^2 w_b}{\partial x_1^2} - M_1^s \frac{\partial^2 w_s}{\partial x_1^2} - M_2^b \frac{\partial^2 w_b}{\partial x_2^2} - M_2^s \frac{\partial^2 w_s}{\partial x_2^2} \\ &\qquad - 2M_{12}^b \frac{\partial^2 w_b}{\partial x_1 \partial x_2} - 2M_{12}^s \frac{\partial^2 w_s}{\partial x_1 \partial x_2} + Q_{13} \frac{\partial w_s}{\partial x_1} + Q_{23} \frac{\partial w_s}{\partial x_2} \Bigg\}, \end{aligned} \tag{2.4.20}$$

where M terms are the horizontal stress resultants defined by:

$$M_1^b = \int_{-h/2}^{h/2} z\sigma_1 \, dz, \qquad M_2^b = \int_{-h/2}^{h/2} z\sigma_2 \, dz, \qquad M_{12}^b = \int_{-h/2}^{h/2} z\sigma_{12} \, dz, \tag{2.4.21}$$

$$M_1^s = \int_{-h/2}^{h/2} \frac{4z^3}{3h^2}\sigma_1 \, dz, \qquad M_2^s = \int_{-h/2}^{h/2} z\frac{4z^3}{3h^2}\sigma_2 \, dz, \qquad M_{12}^s = \int_{-h/2}^{h/2} \frac{4z^3}{3h^2}\sigma_{12} \, dz, \tag{2.4.22}$$

and Q denotes the transversal stress resultants defined as:

$$Q_{13} = \int_{-h/2}^{h/2} \left(1 - \frac{4z^2}{h^2}\right) \sigma_{13} \, dz, \qquad Q_{23} = \int_{-h/2}^{h/2} \left(1 - \frac{4z^2}{h^2}\right) \sigma_{23} \, dz. \tag{2.4.23}$$

Using the constitutive law (2.4.1), one can derive the stress resultants as:

$$\begin{aligned}\begin{bmatrix} M_1^b \\ M_2^b \\ M_{12}^b \end{bmatrix} &= D \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 w_b}{\partial x_1^2} - \tau_c \frac{\partial^2 \dot{w}_b}{\partial x_1^2} \\ -\frac{\partial^2 w_b}{\partial x_2^2} - \tau_c \frac{\partial^2 \dot{w}_b}{\partial x_2^2} \\ -2\frac{\partial^2 w_b}{\partial x_1 \partial x_2} - 2\tau_c \frac{\partial^2 \dot{w}_b}{\partial x_1 \partial x_2} \end{bmatrix} \\ &\quad + cF \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 w_s}{\partial x_1^2} - \tau_c \frac{\partial^2 \dot{w}_s}{\partial x_1^2} \\ -\frac{\partial^2 w_s}{\partial x_2^2} - \tau_c \frac{\partial^2 \dot{w}_s}{\partial x_2^2} \\ -2\frac{\partial^2 w_s}{\partial x_1 \partial x_2} - 2\tau_c \frac{\partial^2 \dot{w}_s}{\partial x_1 \partial x_2} \end{bmatrix}, \end{aligned} \tag{2.4.24}$$

$$\begin{aligned}\begin{bmatrix} M_1^s \\ M_2^s \\ M_{12}^s \end{bmatrix} &= cF \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 w_b}{\partial x_1^2} - \tau_c \frac{\partial^2 \dot{w}_b}{\partial x_1^2} \\ -\frac{\partial^2 w_b}{\partial x_2^2} - \tau_c \frac{\partial^2 \dot{w}_b}{\partial x_2^2} \\ -2\frac{\partial^2 w_b}{\partial x_1 \partial x_2} - 2\tau_c \frac{\partial^2 \dot{w}_b}{\partial x_1 \partial x_2} \end{bmatrix} \\ &\quad + c^2 H \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{bmatrix} -\frac{\partial^2 w_s}{\partial x_1^2} - \tau_c \frac{\partial^2 \dot{w}_s}{\partial x_1^2} \\ -\frac{\partial^2 w_s}{\partial x_2^2} - \tau_c \frac{\partial^2 \dot{w}_s}{\partial x_2^2} \\ -2\frac{\partial^2 w_s}{\partial x_1 \partial x_2} - 2\tau_c \frac{\partial^2 \dot{w}_s}{\partial x_1 \partial x_2} \end{bmatrix}, \end{aligned} \tag{2.4.25}$$

and

$$\begin{bmatrix} Q_{13} \\ Q_{23} \end{bmatrix} = A^s \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{\partial w_s}{\partial x_1} + \tau_c \frac{\partial \dot{w}_s}{\partial x_1} \\ \frac{\partial w_s}{\partial x_2} + \tau_c \frac{\partial \dot{w}_s}{\partial x_2} \end{bmatrix}, \tag{2.4.26}$$

where $c = \frac{4}{3h^2}$ and

$$\begin{aligned} F &:= \int_{-h/2}^{h/2} z^4 \frac{E}{1-\nu^2} dz = \frac{Eh^5}{80(1-\nu^2)}, \\ H &:= \int_{-h/2}^{h/2} z^6 \frac{E}{1-\nu^2} dz = \frac{Eh^7}{448(1-\nu^2)}, \\ A^s &:= \int_{-h/2}^{h/2} \left(1 - \frac{4z^2}{h^2}\right)^2 \frac{E}{2(1+\nu)} dz = \frac{4Eh}{15(1+\nu)}. \end{aligned} \tag{2.4.27}$$

We express the work done by the vertical force of the framework as:

$$\Pi_f(w_b\,, w_s) = -\int_{\Omega_0} \left(q + \eta(\dot{w}_b + \dot{w}_s)\right)(w_b + w_s) \tag{2.4.28}$$

where

$$q = \frac{k_1}{2}(w_b + w_s) - \frac{k_2}{4}(w_b + w_s)^4 + \frac{k_3}{6}(w_b + w_s)^6. \tag{2.4.29}$$

Thus, the system's energy is given by the summation of the potential energy (2.4.20) and the work done by the external force (2.4.28) and the framework response as:

$$\begin{aligned}\Pi(w_b\,,\,w_s) &= \Pi_E(w_b\,,\,w_s) + \Pi_V(w_b\,,\,w_s) + \Pi_f(w_b\,,\,w_s) + \Pi_p(w_b\,,\,w_s)\\ &= \int_{\Omega_0} \mathscr{L}\left(x_1, x_2, \nabla(w_b), \nabla(w_s), \boldsymbol{H}(w_b), \boldsymbol{H}(w_s), \boldsymbol{H}(\dot{w}_b), \boldsymbol{H}(\dot{w}_s)\right).\end{aligned} \tag{2.4.30}$$

Next, we minimize the energy functional (2.4.30) using an Euler-Lagrangian equation. For this functional with two functions and several variables, we use:

$$\begin{aligned}&-\frac{\partial \mathscr{L}}{\partial w_b} + \frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial w_{b,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{b,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial w_{b,11}}\right)\\ &\qquad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{b,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial w_{b,22}}\right)\\ &\qquad + \frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{b,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{b,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{b,11}}\right)\\ &\qquad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{b,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{b,22}}\right) = 0_b,\\ &-\frac{\partial \mathscr{L}}{\partial w_s} + \frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial w_{s,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{s,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial w_{s,11}}\right)\\ &\qquad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial w_{s,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial w_{s,22}}\right)\\ &\qquad + \frac{\partial}{\partial x_1}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{s,1}}\right) + \frac{\partial}{\partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{s,2}}\right) - \frac{\partial^2}{\partial x_1^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{s,11}}\right)\\ &\qquad - \frac{\partial^2}{\partial x_1 \partial x_2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{s,12}}\right) - \frac{\partial^2}{\partial x_2^2}\left(\frac{\partial \mathscr{L}}{\partial \dot{w}_{s,22}}\right) = 0_s,\end{aligned} \tag{2.4.31}$$

Substitution (2.4.30) into (2.4.31), we have:

$$\begin{aligned}&D\Delta^2 w_b + 2\tau_c D\Delta^2 \dot{w}_b + cF\Delta^2 w_s + 2\tau_c cF\Delta^2 \dot{w}_s\\ &+ P\Delta w_b + q + \eta\,(\dot{w}_b + \dot{w}_s) = 0.\\ &cF\Delta^2 w_b + 2\tau_c cF\Delta^2 \dot{w}_b + c^2 H\Delta^2 w_s + 2\tau_c c^2 H\Delta^2 \dot{w}_s - A^s\Delta w_s - 2\tau_c A^s\Delta \dot{w}_s\\ &+ P\Delta w_s + q + \eta\,(\dot{w}_b + \dot{w}_s) = 0,\end{aligned} \tag{2.4.32}$$

Following closely our analysis in 2.3.1, the dimensionless model is obtained as:

$$\begin{aligned}&\Delta^2 w_b + \lambda\Delta^2 \dot{w}_b + \frac{1}{5}\Delta^2 w_s + \frac{1}{5}\lambda\Delta^2 \dot{w}_s + P\Delta w_b + w_b + w_s - (w_b + w_s)^3\\ &+ \mu(w_b + w_s)^5 + \dot{w}_b + \dot{w}_s = 0,\\ &\frac{1}{5}\Delta^2 w_b + \frac{1}{5}\lambda\Delta^2 \dot{w}_b + \frac{1}{21}\Delta^2 w_s + \frac{1}{21}\lambda\Delta^2 \dot{w}_s - A_D^s\Delta w_s - \lambda A_D^s\Delta \dot{w}_s + P\Delta w_s\\ &+ w_b + w_s - (w_b + w_s)^3 + \mu(w_b + w_s)^5 + \dot{w}_b + \dot{w}_s = 0.\end{aligned} \tag{2.4.33}$$

with the following definitions:

$$x_1 \to \left(\frac{D}{k_1}\right)^{1/4} x_1, \qquad x_2 \to \left(\frac{D}{k_1}\right)^{1/4} x_2,$$
$$w_b \to \sqrt{\frac{k_1}{k_2}} w_b, \qquad w_s \to \sqrt{\frac{k_1}{k_2}} w_s,$$
$$P \to \frac{P}{\sqrt{Dk_1}}, \qquad \mu \to \frac{k_1 k_3}{k_2^2},$$
$$A_D^s \to \frac{A_s}{\sqrt{Dk_1}}, \qquad \eta_2 \to \frac{k_1^3}{\eta_1^3 k_2},$$
$$\lambda \to \frac{2\tau_c k_1}{\eta t} \qquad t \to \frac{\eta}{k_1} t.$$

In the next section, we develop a robust numerical simulator for these models.

Remark 4. *We assume that the required boundary conditions and an initial state are given for the models* (2.3.28)*,* (2.4.12)*, and* (2.4.33)*. In Section 2.5, we detail how we choose the corresponding boundary conditions.*

2.5 Numerical procedure

In this section, we describe an efficient simulator for the models we propose (2.3.28), (2.4.12), and (2.4.33). We discretize in space using isogeometric analysis (IGA). IGA is a finite element method where we use NURBS as basis functions, which deliver elements with higher global continuity up to C^{p-1} where p is the polynomial order [28, 29, 38, 99]. This feature allows us to discretise weak formulations with higher-order derivatives (e.g. here, second-order derivatives) [17, 46]. We compare the quadratic functions with continuity of C^0 and C^1 in figure 2.7 to show how the regularity of basis functions affects their shape and support. Later, we use the generalized-α method as a time marching approach to obtain the fully discrete system.

2.5.1 Weak form

Defining a functional space $\mathcal{V} = \mathcal{H}^2$ as a Sobolev space of square-integrable functions with square-integrable first and second derivatives, we multiply models (2.3.28) and (2.4.12) by $v \in \mathcal{V}$. Then, repeating the divergence theorem, the variational formulations are stated as follow, respectively: Find $w \in \mathcal{V}$ such that:

$$\begin{aligned} B(w, v) &= 0, \qquad \forall v \in \mathcal{V}, \\ B_V(w, v) &= 0, \qquad \forall v \in \mathcal{V}, \end{aligned} \tag{2.5.1}$$

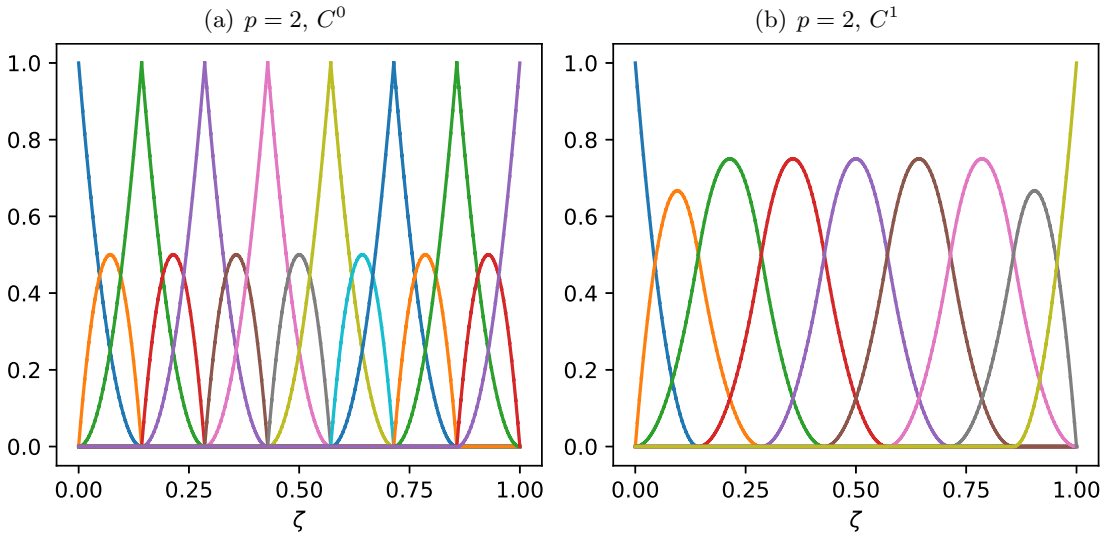


Figure 2.7: Quadratic basis functions for FEM and IGA with C^0 and C^1 continuity, respectively.

where

$$\begin{aligned} B(w,v) &:= (\Delta w, \Delta v) - (P\nabla w, \nabla v) + (\dot{w} + w - w^3 + \mu w^5, v), \\ B_V(w,v) &:= (\Delta w, \Delta v) + (\lambda \Delta \dot{w}, \Delta v) - (P\nabla w, \nabla v) + (\dot{w} + w - w^3 + \mu w^5, v), \end{aligned} \tag{2.5.2}$$

with $(\cdot\,,\,\cdot)$ being the L^2 inner product with respect to the spatial domain. Furthermore, we derive the weak from of the problem (2.4.33) by multiplying it by test functions $v,\, r \in \mathcal{V}$. Then, we obtain the weak problem as finding $w_b,\, w_s \in \mathcal{V}$, such that:

$$B_T(w_b,\, w_s,\, v,\, r) = 0, \qquad \forall\, v,\, r \in \mathcal{V}, \tag{2.5.3}$$

with defining

$$\begin{aligned} B_T(w_b,\, w_s,\, v,\, r) := {} & (\Delta w_b, \Delta v) + (\lambda \Delta \dot{w}_b, \Delta v) + \frac{1}{5}(\Delta w_s, \Delta v) + \frac{1}{5}(\lambda \Delta \dot{w}_s, \Delta v) \\ & - (P\nabla w_b, \nabla v) + \left(w_b + w_s - (w_b + w_s)^3 + \mu(w_b + w_s)^5 + \dot{w}_b + \dot{w}_s,\, v\right), \\ & + \frac{1}{5}(\Delta w_b, \Delta r) + \frac{1}{5}(\lambda \Delta \dot{w}_b, \Delta r) + \frac{1}{21}(\Delta w_s, \Delta r) \\ & + \frac{1}{21}(\lambda \Delta \dot{w}_s, \Delta r) - (P\nabla w_s, \nabla r) \\ & + \left(w_b + w_s - (w_b + w_s)^3 + \mu(w_b + w_s)^5 + \dot{w}_b + \dot{w}_s,\, r\right). \end{aligned} \tag{2.5.4}$$

2.5.2 The semi-discrete formulation

We use a spatial finite element discretization and define $\mathcal{P}_h$ as a partition of the domain Ω into elements K and obtain $\Omega_h := \bigcup_{K\in\mathcal{P}_h} K$. Following standard notation for the Lebesgue and Sobolev spaces, we assume $\mathcal{V}^h$ as a finite-dimensional space composed of polynomial functions with order $p \geq 1$ defined on Ω_h. Then for the spatial discretization of (2.5.1), we use the Galerkin method and approximate (2.5.1) by the following variational problems over the finite-dimensional spaces: find $w^h \in \mathcal{V}^h \subset \mathcal{V}$ such that:

$$\begin{aligned} B(w^h, v^h) &= 0, \qquad \forall v^h \in \mathcal{V}^h, \\ B_N(w^h, v^h) &= 0, \qquad \forall v^h \in \mathcal{V}^h. \end{aligned} \tag{2.5.5}$$

Following the same approach, for (2.5.3), we can state the approximate problem as finding w_b^h, $w_s^h \in \mathcal{V}^h \subset \mathcal{V}$ such that:

$$B_T(w_b^h,\, w_s^h,\, v^h,\, r^h) = 0, \qquad \forall v^h,\, r^h \in \mathcal{V}^h, \tag{2.5.6}$$

with trial solutions w^h, w_s^h, w_s^h defined as:

$$w^h = \sum_{A=1}^{n_b} w_A N_A, \qquad w_b^h = \sum_{A=1}^{n_b} w_{b\,A} N_A, \qquad w_s^h = \sum_{A=1}^{n_b} w_{s\,A} N_A, \tag{2.5.7}$$

and the weighting functions v^h, r^h as:

$$v^h = \sum_{A=1}^{n_b} v_A N_A, \qquad r^h = \sum_{A=1}^{n_b} r_A N_A. \tag{2.5.8}$$

Here, N_A is the basis function, and n_b is the dimension of the discrete space. Based on the definition of $\mathcal{V}_h$, we need the discrete space to be at least $\mathcal{H}^2$-conforming globally. In this work, we satisfy this requirement by using a NURBS basis with C^1-continuity.

2.5.3 Time integration

Let us partition the temporal domain $[0, T]$ uniformly with a time step τ: $0 = t_0 < t_1 < \cdots < t_N = T$ and denote by W_n^h, V_n^h the approximations to $w^h(t_n), \dot{w}^h(t_n)$, respectively. Therefore, exploiting the generalized-α method [60], the fully discrete formulation of the model (2.3.28) at time-step $n+1$ is:

$$\begin{aligned} 0 &= (\Delta W_{n+\alpha_f}^h, \Delta v^h) - (P\nabla W_{n+\alpha_f}^h, \nabla v^h) + \left(V_{n+\alpha_m}^h + \psi(W_{n+\alpha_f}^h), v^h\right), \\ W_{n+1} &= W_n + \tau V_n + \tau\gamma[\![V_n]\!], \\ (V_0^h,\, v^h) &= -(\Delta W_0^h, \Delta v^h) + (P\nabla W_0^h, \nabla v^h) - (\psi(W_0), v^h), \end{aligned} \tag{2.5.9}$$

where

$$\begin{aligned} \psi(W_n) &= W_n^h - \left(W_n^h\right)^3 + \mu \left(W_n^h\right)^5, \\ W_{n+\alpha_f} &= W_n + \alpha_f [\![W_n]\!], \\ V_{n+\alpha_m} &= V_n + \alpha_m [\![V_n]\!], \\ [\![W_n]\!] &= W_{n+1} - W_n. \end{aligned} \tag{2.5.10}$$

According to (2.5.9), the method requires a two-step computation; the first one solves an implicit system to find $[\![V_n]\!]$, the second one uses the second equation in (2.5.9) to update U_{n+1} explicitly. We guarantee the scheme's second-order accuracy in time by setting $\gamma = \frac{1}{2} + \alpha_m - \alpha_f$; we also control the numerical dissipation using the following parameter definition [60]:

$$\alpha_m = \frac{1}{2}\Big(\frac{3-\rho_\infty}{1+\rho_\infty}\Big), \qquad \alpha_f = \frac{1}{1+\rho_\infty}. \tag{2.5.11}$$

where, $\rho_\infty \in [0,1]$ is a user-control parameter. The next chapter describes the dissipation control, a higher-order version of this approximation in time, and the stability behaviour in the temporal domain of the methods we propose.

Remark 5. *To complete our methodology for the implementation of the fully discrete problem* (2.5.9)*, we use Newton's method to solve the resulting nonlinear equation. The iterative procedure can be summarized as computing the residual* $\boldsymbol{R}^{(i)}$*, using* $W_{n+1}^{(i)}$ *with* i *denoting the* i*-th iteration. Next, computing the Jacobian* $\boldsymbol{j}^{(i)}$*, we solve*

$$\boldsymbol{j}^{(i)} \delta W^{(i+1)} = \boldsymbol{R}^i. \tag{2.5.12}$$

Then, we update $W_{n+1}^{(i+1)} = W_{n+1}^{(i)} - \delta W^{(i+1)}$*. We repeat these steps until the norm of the global residual vector is reduced to a certain tolerance (for example, here,* 10^{-8}*) of their initial value. We usually achieve convergence in 3 to 5 nonlinear iterations per time step. For more details, we refer the reader to [46, 100].*

Remark 6. *To obtain the fully discrete version of the models* (2.4.12) *and* (2.4.33) *we follow the procedure of deriving* (2.5.9) *from* (2.5.1) *closely.*

2.5.4 Results and discussion

In our examples, we simulate a plate with an imperfection at its initial state given by:

$$w(\cdot\,,0) = \frac{0.0001}{(1+x^2)(1+y^2)}. \tag{2.5.13}$$

For thick layers and in the case of using (2.4.33), the initial conditions read:

$$w_b(\cdot\,,0) = \frac{0.0001}{(1+x^2)(1+y^2)}, \qquad w_s(\cdot\,,0) = 0. \tag{2.5.14}$$

Then, we study the effects of the viscosity of the layer on the localized buckling by comparing the evolution of folding for different viscosities. For this, we simulate (2.3.28) and (2.4.12) with given $\eta = 0.25$, with a compressive force $P = 4$ and Dirichlet boundary conditions for vertical displacement and the bending moment. Later, we choose time step of $\tau = 0.05$, polynomial space of order $p = 2$ with C^1 continuity. We present the evolution of the fold in figure 2.8. We conclude that the folds approach a sinusoidal shape by decreasing the viscosity of the layer.

We also solve (2.4.12) for a 2D plate and present the result at time 10 in figure 2.9 with cross-sections to show the behaviour of our model. The localized bucklings occur in x and y directions. That is, the sinusoidal patterns predicted by Biot in [14] are not valid in both horizontal dimensions, and the results are more realistic. For example, here, the simulation resembles a dome and basins.

Next, we simulate the model (2.4.33) for a thick layer of thickness $h = 1$. For this, we consider a 2D domain $[-40, 40] \times [-40, 40]$ and set $\mu = 0.25$ with a compressive force $P = 4$. Then, we study the effects of viscosity of the layer on the localized buckling by comparing the evolution of two cases of $\lambda = 0$ and $\lambda = 100$. Figure 2.10 presents the vertical deformation of the layers. The temporal evolution of the folded layer with $\lambda = 100$ results in the formation of patterns. These simulations show that the viscosity of the layer plays an important role in forming localized folds.

In figures 2.11 and 2.12, we show the cross-sections of the elastic and viscoelastic folded plates, respectively. The evolution of the elastic layer shows that the sinusoidal pattern of the deformation remains, while the viscoelastic layers demonstrate varying localized patterns during the simulation time.

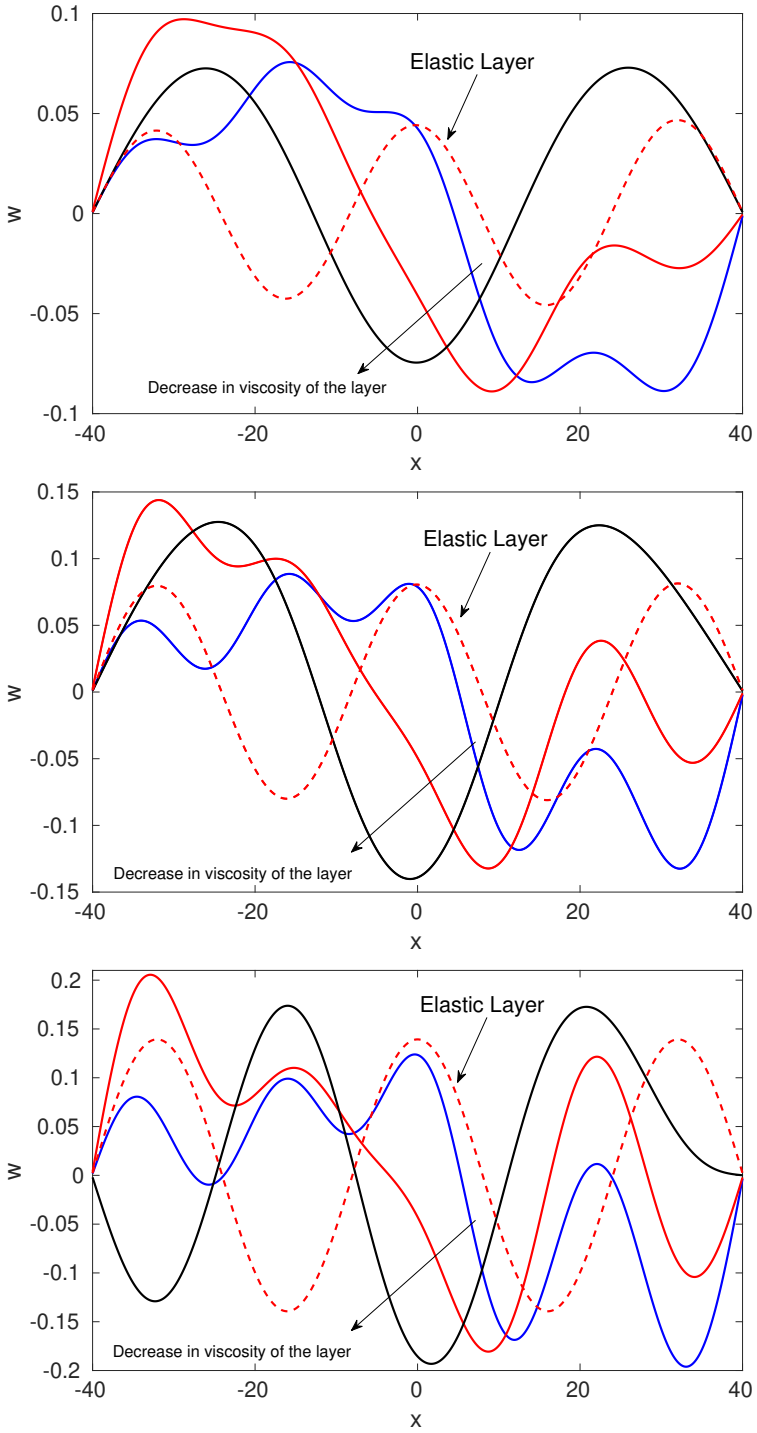


Figure 2.8: From top to bottom: time=3, 10, 30.

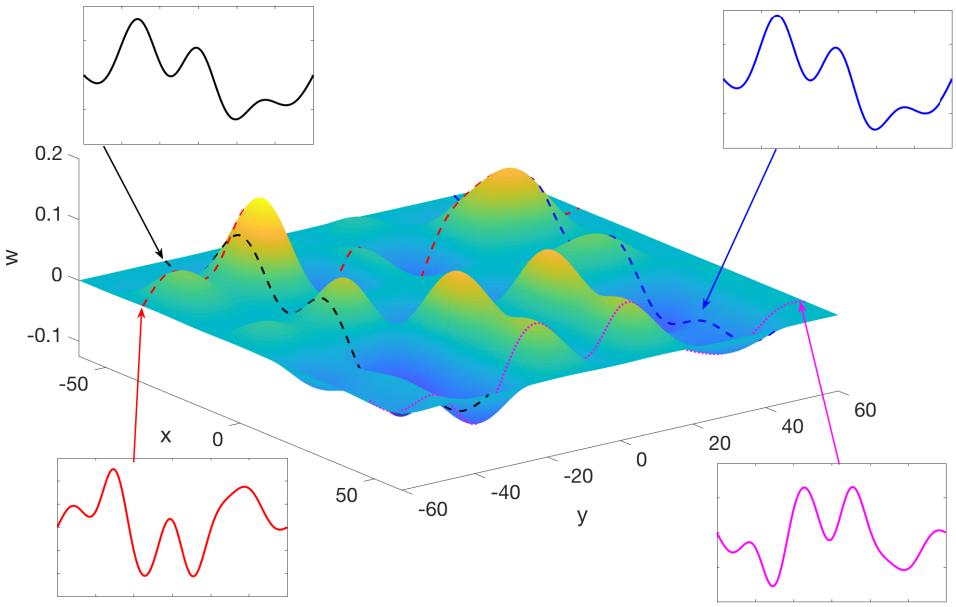


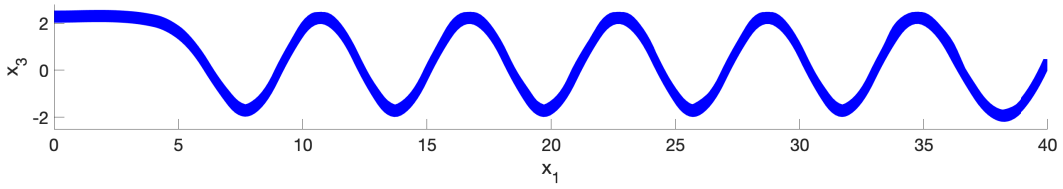
Figure 2.9: The folding of a viscoelastic layer with $\lambda = 1$ at time=10. The introduced model captures the locallized foldins and the structure resembles dome and basins

(a) $\lambda = 0$, $t = 5$ (b) $\lambda = 0$, $t = 100$ (c) $\lambda = 0$, $t = 200$

(d) $\lambda = 100$, $t = 5$ (e) $\lambda = 100$, $t = 100$ (f) $\lambda = 100$, $t = 200$

Figure 2.10: The deformation of the middle surface Ω_0 of the layer embedded in a viscoelastic framework. The localized behaviour is true for all stages of the deformation of the viscoelastic layer.

(a) $\lambda = 0$, $t = 100$



(b) $\lambda = 0$, $t = 200$

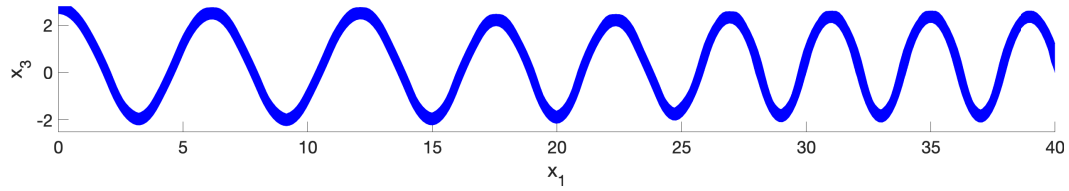
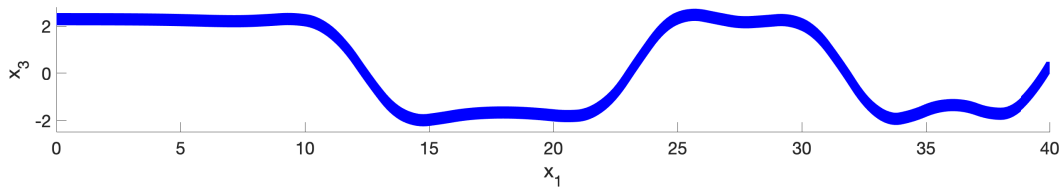


Figure 2.11: The plate's cross-section evolution at $x_2 = 4$ for the case of elastic layer.

(a) $\lambda = 100$, $t = 100$



(b) $\lambda = 100$, $t = 200$

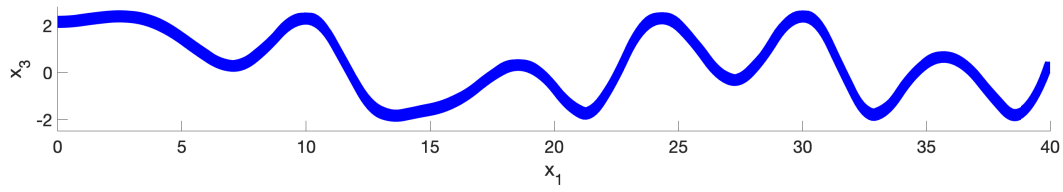


Figure 2.12: The plate's cross-section evolution at $x_2 = 4$.

3

Higher-order Time Marching

Now, we introduce our work on addressing the heretofore-unfulfilled need for higher-order time-integration methods with user-controlled dissipation in the high-frequency part of the spectrum. The commonly-used implicit generalized-α time-marching approach delivers second-order accuracy in time for parabolic and hyperbolic problems. Furthermore, the user can control the numerical dissipation in the high-frequency regions of the discrete spectrum. Nevertheless, we require high-order time marching methods in practice in various cases. For example, to fully exploit the high accuracy of the spatial discretization, it is vital to have a high-order time integrator that handles the possible poor approximability in the discrete high-frequency range. Thus, we extend the generalized-α method to increase its accuracy while keeping its appealing features such as unconditional stability and algorithmic control on high-frequency numerical dissipation. The core idea is to use high-order Taylor's expansion for defining the parameters.

Below, we introduce our approach for the equations with second-order derivative in time (e.g., second-order hyperbolic models) and then approximate the first-order derivative in time (e.g., parabolic models). Then, we introduce a new class of explicit time integrators delivering an accuracy order of $2k$, $k \in \mathbb{N}$ as well as the user-control on the numerical dissipations. We also detail our analysis on the unconditional stability of implicit methods, the CFL conditions for the explicit technique, and the accuracy of the methods.

For this chapter, we use the methodologies and the results that we reported in the following papers:

- Behnoudfar, Pouria, Quanling Deng, and Victor M. Calo. "Higher-order generalized-α methods for hyperbolic problems. "Computer Methods in Applied Mechanics and Engineering 378 (2021): 113725 [10].
- Behnoudfar, Pouria, Quanling Deng, and Victor M. Calo. "High-order generalized-α method." Applications in Engineering Science 4 (2020): 100021 [8].
- Behnoudfar, Pouria, Gabriele Loli, Alessandro Reali, Giancarlo Sangalli, and Victor M. Calo. "Explicit high-order generalized-α methods for isogeometric analysis of structural dynamics." arXiv preprint arXiv:2102.11536 (2021) [12].
- Behnoudfar, Pouria, Quanling Deng, and Victor M. Calo. "Higher-order generalized-α methods for parabolic problems." arXiv preprint arXiv:2102.05910 (2021) [9].

In the next sections, we summarise our schemes as well as their analysis and use.

3.1 Implicit higher-order time marching methods for hyperbolic problems

3.1.1 Generalized-α for second-order time derivatives

We first consider a second-order ordinary differential equation (ODE)

$$\begin{aligned} \ddot{u} + \lambda u &= 0, \\ u(0) &= u_0, \\ \dot{u}(0) &= v_0, \end{aligned} \tag{3.1.1}$$

with the given initial values of the solution, u_0, and v_0 its first derivative. We discretise the time-marching domains between $0 = t_0 < t_1 < \cdots < t_N = T$ with T being the final time and introducing the time step-size as $\tau_n = t_{n+1} - t_n$. Herein, without loss of generality, we consider a uniform grid with time step τ. Denoting the approximations of $u(t_n)$, $\dot{u}(t_n)$, $\ddot{u}(t_n)$, by U_n, V_n, A_n respectively, the generalized-α method for solving (3.1.1) at time-step $n+1$ is given by [26]:

$$A_{n+\alpha_m} = -\lambda U_{n+\alpha_f}, \tag{3.1.2a}$$

$$U_{n+1} = U_n + \tau v_n + \frac{\tau^2}{2} A_n + \tau^2 \beta_1 [\![A_n]\!], \tag{3.1.2b}$$

$$V_{n+1} = V_n + \tau A_n + \tau \gamma_1 [\![A_n]\!], \tag{3.1.2c}$$

with the discrete counterparts of the given initial data $U_0 = u_0$ and $V_0 = v_0$, and the estimated initial acceleration $A_0 = -\lambda U_0$, and the parameters defined using (2.5.10). Setting $\gamma_1 = \frac{1}{2} + \alpha_m - \alpha_f$ guarantees second-order of accuracy. To control the numerical dissipation, we set $\beta_1 = \frac{1}{4}(1 + \alpha_m - \alpha_f)^2$ and follow the standard parameter definitions:

$$\alpha_f = \frac{1}{1+\rho_\infty}, \qquad \alpha_m = \frac{2-\rho_\infty}{1+\rho_\infty}, \tag{3.1.3}$$

where $\rho_\infty \in [0, 1]$ is a control parameter that the user can adjust to control the high-frequency dissipation of the methods.

3.1.2 A fourth-order generalized-α method

Equations (3.1.2b)-(3.1.2c) imply a sub-step time-marching for the generalized-α scheme that limit the order of accuracy as the truncation error is of order $\mathcal{O}(\tau^3)$. To overcome this limitation, we exploit a higher-order Taylor expansion to obtain these two equations; see our detailed discussion in [8]. For this, let $\mathcal{L}^a(s)$ be the a-th order derivative of the function s in time. Therefore, for example, to formulate a fourth-order generalized-α method, we introduce a method that solves

$$\begin{aligned} A_n^{\alpha_1} &= -\lambda U_{n+1}, \\ \mathcal{L}^3(A_n^{\alpha_2}) &= -\lambda \mathcal{L}^1(A_n^{\alpha_f}), \end{aligned} \tag{3.1.4}$$

with the following update strategy

$$\begin{aligned} U_{n+1} &= U_n + \tau V_n + \frac{\tau^2}{2} A_n + \frac{\tau^3}{6}\mathcal{L}^1(A_n) + \frac{\tau^4}{24}\mathcal{L}^2(A_n) + \frac{\tau^5}{120}\mathcal{L}^3(A_n) + \beta_1 \tau^2 P_n, \\ V_{n+1} &= V_n + \tau A_n + \frac{\tau^2}{2}\mathcal{L}^1(A_n) + \frac{\tau^3}{6}\mathcal{L}^2(A_n) + \frac{\tau^4}{24}\mathcal{L}^3(A_n) + \gamma_1 \tau P_n, \end{aligned}$$

$$\begin{aligned} \mathcal{L}^1(A_{n+1}) &= \mathcal{L}^1(A_n) + \tau\mathcal{L}^2(A_n) + \frac{\tau^2}{2}\mathcal{L}^3(A_n) + \tau^2\beta_2[\![\mathcal{L}^3(A_n)]\!], \\ \mathcal{L}^2(A_{n+1}) &= \mathcal{L}^2(A_n) + \tau\mathcal{L}^3(A_n) + \tau\gamma_2[\![\mathcal{L}^3(A_n)]\!], \end{aligned} \tag{3.1.5}$$

where

$$\begin{aligned} P_n &= A_{n+1} - A_n - \tau\mathcal{L}^1(A_n) - \frac{\tau^2}{2}\mathcal{L}^2(A_n) - \frac{\tau^3}{6}\mathcal{L}^3(A_n), \\ A_n^{\alpha_1} &= A_n + \tau\mathcal{L}^1(A_n) + \frac{\tau^2}{2}\mathcal{L}^2(A_n) + \frac{\tau^3}{6}\mathcal{L}^3(A_n) + \alpha_1 P_n, \\ \mathcal{L}^3(A_n^{\alpha_2}) &= \mathcal{L}^3(A_n) + \alpha_2[\![\mathcal{L}^3(A_n)]\!], \\ \mathcal{L}^1(A_n^{\alpha_f}) &= \mathcal{L}^1(A_n) + \alpha_f[\![\mathcal{L}^1(A_n)]\!]. \end{aligned} \tag{3.1.6}$$

Assuming sufficient regularity for the solution in time, one can readily obtain $\mathcal{L}^3(A_n^{\alpha_2}) = -\lambda\mathcal{L}^1(A_n^{\alpha_f})$ by taking three derivatives of the first equation in (3.1.4) with respect to time. Also, using the given information U_0 and V_0, we determine:

$$\begin{aligned} A_0 = -\lambda U_0, \qquad \mathcal{L}^1(A_0) &= -\lambda V_0, \\ \mathcal{L}^2(A_0) = \lambda^2 U_0, \qquad \mathcal{L}^3(A_0) &= \lambda^2 V_0, \end{aligned} \tag{3.1.7}$$

which completes a brief introduction of our method. Next, we now determine the accuracy order and stability of the method.

3.1.2.1 Order of accuracy in time

As an example, we now derive the fourth-order accuracy of (3.1.4), we require to set conditions on the parameters γ_1 and γ_2 . For this, from (3.1.5) and (3.1.4), we obtain a system of equations for each time step as

$$A\mathbf{U_{n+1}} = B\mathbf{U_n}, \tag{3.1.8}$$

where

$$A = \begin{bmatrix} 1 & 0 & -\beta_1 & 0 & 0 & 0 \\ 0 & 1 & -\gamma_1 & 0 & 0 & 0 \\ \tau^2\lambda & 0 & \alpha_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -\beta_2 \\ 0 & 0 & 0 & 0 & 1 & -\gamma_2 \\ 0 & 0 & 0 & \tau^2\alpha_f\lambda & 0 & \alpha_2 \end{bmatrix}, \qquad \mathbf{U_n} = \begin{bmatrix} U_n \\ \tau V_n \\ \tau^2 A_n \\ \tau^3\mathcal{L}^1(A_n) \\ \tau^4\mathcal{L}^2(A_n) \\ \tau^5\mathcal{L}^3(A_n) \end{bmatrix},$$

$$B = \frac{1}{120}\begin{bmatrix} 120 & 120 & 60\,(1-2\,\beta_1) & 20\,(1-6\,\beta_1) & 5\,(1-12\,\beta_1) & 1-20\,\beta_1 \\ 0 & 120 & 120\,(1-\gamma_1) & 60\,(1-2\,\beta_1) & 20\,(1-3\,\beta_1) & 5\,(1-4\,\beta_1) \\ 0 & 0 & 120\,(\alpha_1-1) & 120\,(\alpha_2-1) & 60\,(\alpha_2-1) & 20\,(\alpha_2-1) \\ 0 & 0 & 0 & 120 & 120 & 60\,(1-2\,\beta_2) \\ 0 & 0 & 0 & 0 & 120 & 120\,(1-\gamma_2) \\ 0 & 0 & 0 & -120\,\tau^2(1-\alpha_f)\lambda & 0 & 120\,(\alpha_2-1) \end{bmatrix}. \tag{3.1.9}$$

Thus, we define the amplification matrix G as:

$$G = A^{-1}B. \tag{3.1.10}$$

Then, for any arbitrary amplification matrix, we can readily obtain the following [12, 56]:

$$\mathcal{L}^1(A_{n+1}) - (\text{tr } G)\mathcal{L}^1(A_n) + \frac{1}{2}\left((\text{tr } G)^2 - \text{tr } (G^2)\right)\mathcal{L}^1(A_{n-1}) - \det(G)\,\mathcal{L}^1(A_{n-2}) = 0, \tag{3.1.11}$$

with the coefficients being the invariants of G and tr (G) is the trace of G. Next, we estimate the unknowns $\mathcal{L}^1(A)$ and U at time steps $n-2$, $n-1$, and $n+1$ as:

$$\begin{aligned}
\mathcal{L}^1(A_{n+1}) &= \mathcal{L}^1(A_n) + \tau\mathcal{L}^2(A_n) + \frac{\tau^2}{2}\mathcal{L}^3(A_n) + \mathcal{O}(\tau^3),\\
\mathcal{L}^1(A_{n-1}) &= \mathcal{L}^1(A_n) - \tau\mathcal{L}^2(A_n) + \frac{\tau^2}{2}\mathcal{L}^3(A_n) + \mathcal{O}(\tau^3),\\
\mathcal{L}^1(A_{n-2}) &= \mathcal{L}^1(A_n) - 2\tau\mathcal{L}^2(A_n) + 2\tau^2\mathcal{L}^3(A_n) + \mathcal{O}(\tau^3),\\
U_{n+1} &= U_n + \tau V_n + \frac{\tau^2}{2}A_n + \mathcal{R}_1,\\
U_{n-1} &= U_n - \tau V_n + \frac{\tau^2}{2}A_n + \mathcal{R}_2,\\
U_{n-2} &= U_n - 2\tau V_n + 4\frac{\tau^2}{2}A_n + \mathcal{R}_3,
\end{aligned} \tag{3.1.12}$$

with the terms $\mathcal{R}_k$ being defined using a Taylor's expansion as a function of $\tau^3\mathcal{L}^1(A_n)$, $\tau^4\mathcal{L}^2(A_n)$, and $\tau^5\mathcal{L}^3(A_n)$. For example, we have

$$\mathcal{R}_1 = \frac{\tau^3}{6}\mathcal{L}^1(A_n) + \frac{\tau^4}{24}\mathcal{L}^2(A_n) + \frac{\tau^5}{120}\mathcal{L}^3(A_n). \tag{3.1.13}$$

Finally, substituting (3.1.12) into (3.1.11), to have fourth-order of accuracy in time or truncation error of $\mathcal{O}(\tau^5)$, we propose:

$$\gamma_1 = \alpha_1 - \frac{1}{2}, \qquad \gamma_2 = \frac{1}{2} - \alpha_f + \alpha_2. \tag{3.1.14}$$

For more details, we refer the reader to [10].

3.1.2.2 Stability analysis and eigenvalue control

We construct an unconditionally stable method; to do so, we need to bound absolute values of the eigenvalues of the amplification matrix by one. For this, firstly , we determine the eigenvalues of of (3.1.10) for the case $\sigma := \lambda\tau^2 \to 0$ as

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1, \qquad \lambda_5 = \frac{\alpha_1 - 1}{\alpha_1}, \qquad \lambda_6 = \frac{\alpha_2 - 1}{\alpha_2}. \tag{3.1.15}$$

Therefore, the conditions $\alpha_1 \geq \frac{1}{2}$ and $\alpha_2 \geq \frac{1}{2}$ imply the boundedness of λ_5 and λ_6. Similarly, for the case of $\sigma \to \infty$, we propose the following conditions:

$$1 \leq \alpha_1, \qquad \frac{1}{2} \leq \alpha_f \leq \alpha_2. \tag{3.1.16}$$

We also show the stability regions for two cases in which α_1 is constant as well as when

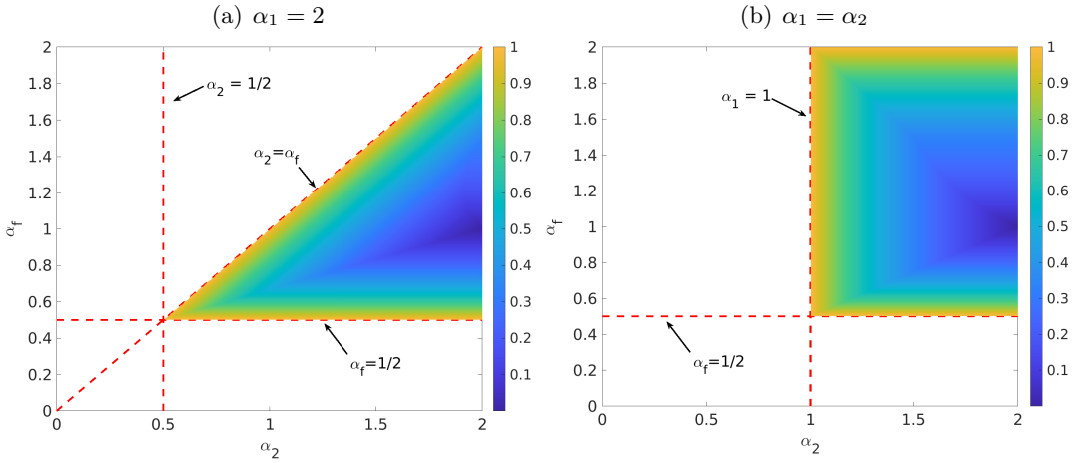


Figure 3.1: Stability regions for (a) $\alpha_1 = 2$, (b) $\alpha_1 = \alpha_2$.

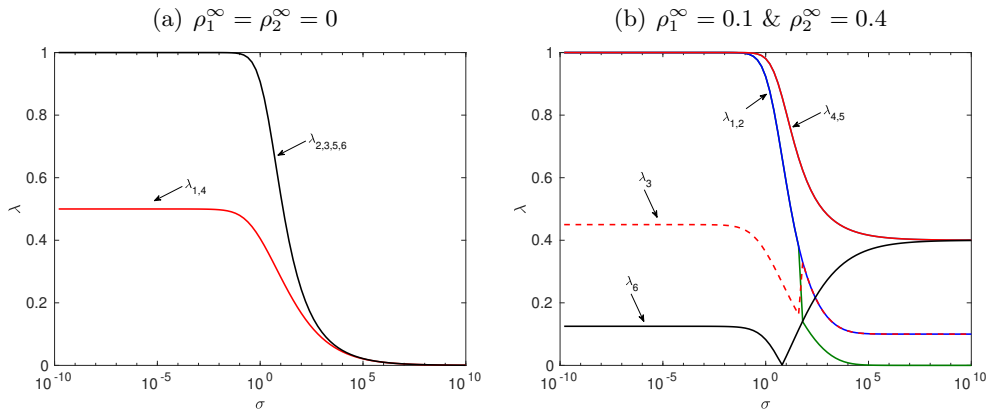


Figure 3.2: Eigenvalues of the amplification matrix, fourth-order-accurate time-marching scheme: (a) $\rho_1^\infty = \rho_2^\infty = 0$, (b) $\rho_1^\infty = 0.1$ & $\rho_2^\infty = 0.4$ [10].

it is equal to α_2 in figure 3.1.

In the next step, to control the numerical dissipation, we follow the analysis of the second-order generalized-α method in [7, 26, 60] and define:

$$\beta_1 = \frac{1}{16}\left(1 + 4\gamma_1 + 4\gamma_1^2\right), \qquad \beta_2 = \frac{1}{16}\left(1 + 4\gamma_2 + 4\gamma_2^2\right). \tag{3.1.17}$$

Finally, we set the user-control parameters ρ_1^∞, ρ_2^∞ equal to the spectral radius of the system and introduce:

$$\alpha_1 = \frac{2}{1+\rho_1^\infty}, \qquad \alpha_f = \frac{1}{1+\rho_2^\infty}, \qquad \alpha_2 = \frac{2-\rho_2^\infty}{1+\rho_2^\infty}. \tag{3.1.18}$$

Therefore, by choosing $0 \leq \rho_1^\infty, \rho_2^\infty \leq 1$, one controls the eigenvalues of the amplification matrix and consequently, the high-frequency damping; see figure 3.2.

Next, we extend our method to an arbitrary order of accuracy $2k$, $k \in \mathbb{N}$ in the temporal domain.

3.1.3 Higher-order accuracy in time

Our scheme delivers the $2k$-th order generalized-α method by solving:

$$\begin{aligned} A_n^{\alpha_1} &= -\lambda U_{n+1}, \\ \mathcal{L}^{3j-3}(A_n^{\alpha_j}) &= -\lambda \mathcal{L}^{3j-5}(A_{n+1}), \qquad j = 2, \cdots, k-1, \\ \mathcal{L}^{3k-3}(A_n^{\alpha_k}) &= -\lambda \mathcal{L}^{3k-5}(A_n^{\alpha_f}), \end{aligned} \tag{3.1.19}$$

and updating the system as follows

$$\begin{aligned} U_{n+1} &= U_n + \tau V_n + \frac{\tau^2}{2} A_n + \frac{\tau^3}{6}\mathcal{L}^1(A_n) + \cdots + \frac{\tau^{3k-1}}{(3k-1)!}\mathcal{L}^{3k-3}(A_n) + \tau^2 \beta_1 P_{n,1}, \\ V_{n+1} &= V_n + \tau A_n + \frac{\tau^2}{2}\mathcal{L}^1(A_n) + \cdots + \frac{\tau^{3k-2}}{(3k-2)!}\mathcal{L}^{3k-3}(A_n) + \tau \gamma_1 P_{n,1}, \\ \mathcal{L}^{3j-5}(A_{n+1}) &= \mathcal{L}^{3j-5}(A_n) + \tau \mathcal{L}^{3j-4}(A_n) + \cdots + \frac{\tau^{3(k-j)+2}}{(3(k-j)+2)!}\mathcal{L}^{3k-3}(A_n) + \tau^2 \beta_j P_{n,j}, \\ \mathcal{L}^{3j-4}(A_{n+1}) &= \mathcal{L}^{3j-4}(A_n) + \cdots + \frac{\tau^{3(k-j)+1}}{(3(k-j)+1)!} + \tau \gamma_j P_{n,j}, \qquad j = 2, \cdots, k-1, \\ \mathcal{L}^{3k-5}(A_{n+1}) &= \mathcal{L}^{3k-5}(A_n) + \tau \mathcal{L}^{3k-4}(A_n) + \frac{\tau^2}{2}\mathcal{L}^{3k-3}(A_n) + \tau^2 \beta_k [\![\mathcal{L}^{3k-3}(A_n)]\!], \\ \mathcal{L}^{3k-4}(A_{n+1}) &= \mathcal{L}^{3k-4}(A_n) + \tau \mathcal{L}^{3k-3}(A_n) + \tau \gamma_k [\![\mathcal{L}^{3k-3}(A_n)]\!], \end{aligned} \tag{3.1.20}$$

where we have

$$\begin{aligned} P_{n,1} &= A_{n+1} - A_n - \tau \mathcal{L}^1(A_n) - \cdots - \frac{\tau^{3k-3}}{(3k-3)!}\mathcal{L}^{3k-3}(A_n), \\ A_n^{\alpha_1} &= A_n + \tau \mathcal{L}^1(A_n) + \cdots + \frac{\tau^{3k-3}}{(3k-3)!}\mathcal{L}^{3k-3}(A_n) + \alpha_1 P_{n,1}, \\ P_{n,j} &= \mathcal{L}^{3j-3}(A_{n+1}) - \mathcal{L}^{3j-3}(A_n) - \cdots - \frac{\tau^{3(k-j)}}{3(k-j)!}\mathcal{L}^{3k-3}(A_n), \\ A_n^{\alpha_j} &= \mathcal{L}^{3j-3}(A_n) + \cdots + \frac{\tau^{3(k-j)}}{3(k-j)!}\mathcal{L}^{3k-3}(A_n) + \alpha_j P_{n,j}, \qquad j = 2, \cdots, k-1, \\ \mathcal{L}^{3k-3}(A_n^{\alpha_k}) &= \mathcal{L}^{3k-3}(A_n) + \alpha_k [\![\mathcal{L}^{3k-3}(A_n]\!], \\ \mathcal{L}^{3k-5}(A_n^{\alpha_f}) &= \mathcal{L}^{k3-5}(A_n) + \alpha_f [\![\mathcal{L}^{3k-5}(A_n]\!]. \end{aligned} \tag{3.1.21}$$

For $k = 1, 2$, one retrieves the second- and fourth-order generalized-α methods, respectively. For the analysis, we a similar approach to the one we discussed in the previous section; therefore, we obtain the matrix system as:

$$\boldsymbol{L}\boldsymbol{U}_{n+1} = \boldsymbol{R}\boldsymbol{U}_n. \tag{3.1.22}$$

Accordingly, the amplification matrix $H = L^{-1}R$ is defined as

$$H = \begin{bmatrix} \Lambda_1 & \Xi_{12} & \cdots & \cdots & \Xi_{1k} \\ 0 & \Lambda_2 & \Xi_{23} & \cdots & \Xi_{2k} \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \Lambda_{k-1} & \Xi_{(k-1)k} \\ 0 & 0 & \cdots & 0 & \Lambda_k \end{bmatrix}, \tag{3.1.23}$$

where

$$\Lambda_i = \sigma_i \begin{bmatrix} \alpha_i & \alpha_i & \frac{1}{2}\alpha_i - \beta_i \\ -\gamma_i\sigma & (\beta_i - \gamma_i)\sigma + \alpha_i & (\beta_i - \frac{1}{2}\gamma_i)\sigma + \alpha_i - \gamma_i \\ -\sigma & -\sigma & (\beta_i - \frac{1}{2})\sigma + \alpha_i - 1 \end{bmatrix}, \quad i = 1, \cdots, k-1,$$

$$\Lambda_k = \sigma_k \begin{bmatrix} \beta_k(\alpha_f - 1)\sigma + \alpha_k & \alpha_k & \frac{1}{2}\alpha_k - \beta_k \\ -\gamma_k\sigma & \alpha_f(\beta_k - \gamma_k)\sigma + \alpha_k & \alpha_f\sigma(\beta_k - \frac{1}{2}\gamma_k) + \alpha_k - \gamma_k \\ -\sigma & -\alpha_f\sigma & \alpha_f\sigma(\beta_k - \frac{1}{2}) + \alpha_k - 1 \end{bmatrix}, \tag{3.1.24}$$

where $\sigma_i = (\beta_i\sigma + \alpha_i)^{-1}$ and $\sigma_k = (\beta_k\alpha_f\sigma + \alpha_k)^{-1}$. We propose to obtain $2k$-order accuracy in time:

$$\begin{aligned} \gamma_i &= \alpha_i - \frac{1}{2}, \qquad \text{for } i = 1, \cdots, k-1, \\ \gamma_k &= \frac{1}{2} - \alpha_f + \alpha_k. \end{aligned} \tag{3.1.25}$$

Due to the structure of the amplification matrix (diagonal block entries), the analysis is similar to what we provided for the fourth-order method. We repeat the approach k times for a scheme with a $2k$ order of accuracy. Here, for brevity, we omit the details and refer the reader to [10].

3.1.3.1 Stability analysis and control on dissipation

To analyze the spectral behaviour of the amplification matrix (3.1.22), we refer to the sketch that figure 3.3 illustrates and (3.1.23), where we calculate all the eigenvalues

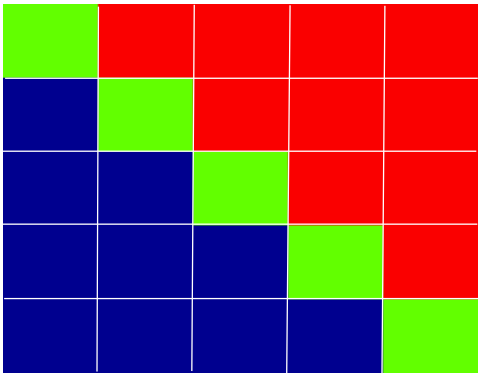


Figure 3.3: $2k$-order accurate method in time: Amplification matrix. Blue blocks have zero entries, and the entries of the green ones are similar to the second-order generalized-α method [10].

by determining the eigenvalues of each diagonal block. Hence, we set the parameters as [10] (following Section 3.1.2.2)

$$\beta_i = \frac{1}{16}\big(1 + 4\gamma_i + 4\gamma_i^2\big), \qquad i = 1, \cdots, k. \tag{3.1.26}$$

Accordingly, we define user-control parameters ρ_∞^i and set:

$$\begin{aligned} \alpha_i &= \frac{2}{1+\rho_i^\infty}, & i &= 1, \cdots, k-1, \\ \alpha_k &= \frac{2-\rho_k^\infty}{1+\rho_k^\infty}, & \alpha_f &= \frac{1}{1+\rho_k^\infty}. \end{aligned} \tag{3.1.27}$$

where the eigenvalues of block i approaches to ρ_i^∞ when $\sigma \to \infty$.

3.1.4 Linear structural dynamics

In this section, we use our time marching scheme to simulate time-dependent PDEs. We consider the initial boundary-value hyperbolic problem of structural dynamics, that is:

$$\begin{aligned} \ddot{u}(\boldsymbol{x},t) - \nabla\cdot\left(\omega^2\,\nabla u(\boldsymbol{x},t)\right) + c\left(\dot{u}(\boldsymbol{x},t)\right) &= f(\boldsymbol{x},t), & (\boldsymbol{x},t) &\in \Omega\times[0,T], \\ u(\boldsymbol{x},t) &= u_D, & \boldsymbol{x} &\in \partial\Omega\times[0,T], \\ u(\boldsymbol{x},0) &= u_0, & \boldsymbol{x} &\in \Omega, \\ \dot{u}(\boldsymbol{x},0) &= v_0, & \boldsymbol{x} &\in \Omega, \end{aligned} \tag{3.1.28}$$

Let $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$, be an open bounded domain with Lipschitz boundary $\partial\Omega$. ∇. The operator ∇ is the spatial gradient while a superscript dot denotes a time derivative

$$\dot{u}(\boldsymbol{x}, t) = \frac{\partial u(\boldsymbol{x}, t)}{\partial t} \text{ and } \ddot{u}(\boldsymbol{x}, t) = \frac{\partial^2 u(\boldsymbol{x}, t)}{\partial t^2}.$$

$c\left(\dot{u}(\boldsymbol{x}, t)\right)$ models linear damping. The source f, propagation speed ω, initial data u_0, v_0, and Dirichlet boundary conditions u_D are given and assumed regular enough for the problem to admit a weak solution. We derive our numerical method for (3.1.28) by first obtaining a semi-discrete problem by using finite elements in space; then, we use our generalized-α method to discretise the problem fully.

Remark 7. *In the considered problem* (3.1.28) *and accordingly* (3.1.32)*, we only consider constant ω and assume that the solution $u(\cdot\,, t)$ satisfies homogeneous boundary condition. One requires slight modifications of the discrete bilinear and linear functions for the cases of heterogeneous propagation speed and non-homogeneous boundary conditions; for more details, see [56].*

We use a spatial finite element discretization. For this, We define $\mathcal{P}_h$ as a partition of the domain Ω into elements K and obtain $\Omega_h := \bigcup_{K\in\mathcal{P}_h} K$. Following standard notation for the Lebesgue and Sobolev spaces, we consider V_h^p as a finite-dimensional space composed of polynomial functions with order $p \geq 1$ defined on Ω_h. Next, we multiply the hyperbolic equation (3.1.28) with a sufficiently regular test function $w_h \in V_h^p$ (here, $w_h \in H_0^1(\Omega)$), integrate over Ω_h, and exploit the divergence theorem to obtain the semi-discrete form of the problem as:

$$(w_h, \ddot{u}_h) + a(w_h, \dot{u}_h) + b(w_h, u_h) = \ell(w_h), \qquad w_h \in V_h^p(\Omega),\ t > 0, \tag{3.1.29}$$

where we have

$$\begin{aligned} a(w, v) &= (w, c(v)), \\ b(w, v) &= (\omega^2 \nabla w, \nabla v), \\ \ell(w) &= (w, f), \end{aligned} \tag{3.1.30}$$

with $(\cdot\,, \cdot)$ being the L_2-inner product defined on the domain Ω_h. We approximate $u(x, t)$ for each fixed t by a function $u_h(x, t)$ that belongs to a finite-dimensional space V_h^p. This allows us to derive the following variational formulation:

$$\begin{cases} \text{Find } u_h(t) = u_h(\cdot, t) \in V_h^p(\Omega) & \text{for } t > 0 : \\ (w_h, \ddot{u}_h) + a(w_h, \dot{u}_h) + b(w_h, u_h) = \ell(w_h), & \forall w_h \in V_h^p \end{cases} \tag{3.1.31}$$

with $u_h(0)$ being the interpolation of u_0 in V^h. Next, adopting a Galerkin method (in particular, isogeometric analysis [28, 29]), the matrix problem resulting from the semi-discretization of the linear structural dynamics problem (3.1.31) reads:

$$M\ddot{U} + C\dot{U} + KU = F, \tag{3.1.32}$$

with M, C and K denoting the mass, the viscous damping, and the stiffness matrices, respectively. F is the given external-force vector, and U is the unknown displacement vector. We assume M and K are symmetric and positive definite. Any spatial discretization methods such as finite differences, finite elements, or Isogeometric analysis that provides such properties can result in the matrix equation (3.1.32). To simulate the initial-value problem (3.1.32), we need to determine a time-dependent function $U(t)$ that satisfies (3.1.32) as well as the given initial conditions:

$$U(0) = U_0, \qquad \dot{U}(0) = V_0. \tag{3.1.33}$$

Then, exploiting our high-order method, to have $2k$ order of accuracy in the temporal domain, the full-discrete problem reads:

$$\begin{aligned}
MA_n^{\alpha_1} + CV_{n+1} + KU_{n+1} &= F_{n+1}, \\
M\mathcal{L}^{3j-3}(A_n^{\alpha_j}) + C\mathcal{L}^{3j-4}(A_{n+1}) + K\mathcal{L}^{3j-5}(A_{n+1}) &= F_{n+1}^{(3j-3)}, \qquad j = 2, \cdots, k-1, \\
M\mathcal{L}^{3k-3}(A_n^{\alpha_k}) + C\mathcal{L}^{3k-4}(A_n^{\alpha_f}) + K\mathcal{L}^{3k-5}(A_n^{\alpha_f}) &= F_{n+\alpha_f}^{(3k-3)},
\end{aligned} \tag{3.1.34}$$

updating as in (3.1.20). We also approximate $\frac{\partial^{(a)}}{\partial t^{(a)}} f(\cdot\,, n)$ as $F_n^{(a)}$, and accordingly, define $F_{n+\alpha_f}^{(3k-3)} = F_n^{(3k-3)} + \alpha_f [\![F_n^{(3k-3)}]\!]$. Additionally, the definitions of other terms can be found in (3.1.21). Lastly, we require to determine the values of $3k-2$ high-order unknowns at the initial step from the given U_0 and V_0. For this, we have:

$$\begin{aligned}
A_0 &= M^{-1}\left[F_0 - CV_0 - KU_0\right], \\
\mathcal{L}^1(A_0) &= M^{-1}\left[F_0^{(1)} - CA_0 - KV_0\right], \\
\mathcal{L}^j(A_0) &= M^{-1}\left[F_0^{(j)} - C\mathcal{L}^{j-1}(A_0) - K\mathcal{L}^{j-2}(A_0)\right], \qquad j = 2, \cdots, 3k-3.
\end{aligned} \tag{3.1.35}$$

Remark 8. *We introduce a new higher-order generalized-α method by considering a spatial discretization, leading to the matrix problem* (3.1.32)*. Therefore, our method and its corresponding analysis are independent of the use of the spatial discretization [10, 12].*

3.1.4.1 Analysis

First, we rewrite the method in a0 matrix form as:

$$L\boldsymbol{U}_{n+1} = R\boldsymbol{U}_n + \boldsymbol{F}, \tag{3.1.36}$$

with $\boldsymbol{F}$ denoting the forcing vector. Without loss of generality, we can assume $\boldsymbol{F} = \boldsymbol{0}$ (for details, see, [26]). Then, we derive the amplitude matrix in a diagonal block structure (see, figure 3.3) similar to (3.1.23). Therefore, to analyse the stability behaviour of the method, we require to determine the eigenvalues of each diagonal block. Starting with the lowest diagonal block Λ_k corresponding to three highest-order terms, we define ω and κ as the largest eigenvalues of the matrices $M^{-1}K$ and $M^{-1}C$, respectively, which allows us to state:

$$\begin{aligned}
\Lambda_k^{11} &= \frac{\alpha_f\gamma_k L + \alpha_k + \beta_k(\alpha_f - 1)S}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S}, & \Lambda_k^{12} &= \frac{-\beta_k L + \alpha_f\gamma_k L + \alpha_k}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S},\\
\Lambda_k^{13} &= \frac{\alpha_f\gamma_k L - 2\beta_k(\alpha_f L + 1) + \alpha_k}{2(\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S)}, & \Lambda_k^{21} &= -\frac{\gamma_k S}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S},\\
\Lambda_k^{22} &= \frac{(\alpha_f - 1)\gamma_k L + \alpha_k + \alpha_f(\beta_k - \gamma_k)S}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S}, & \Lambda_k^{23} &= \frac{2(\alpha_k + \beta_k\alpha_f S) - \gamma_k(\alpha_f S + 2)}{2(\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S)},\\
\Lambda_k^{31} &= -\frac{S}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S}, & \Lambda_k^{32} &= -\frac{L + \alpha_f S}{\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S},\\
\Lambda_k^{33} &= \frac{2(\alpha_f(\gamma_k - 1)L + \alpha_k - 1) + (2\beta_k - 1)\alpha_f S}{2(\alpha_f\gamma_k L + \alpha_k + \beta_k\alpha_f S)},
\end{aligned} \tag{3.1.37}$$

where Λ_k^{ij} is the entry of the matrix Λ_k, $S = \omega\tau^2$, and $L = \kappa\tau$. The entries (3.1.37) leads to an identical to the one obtained for the problem (3.1.1) for the limit cases $\tau \to 0$ and $\tau \to +\infty$. Therefore, our proposed parameters such as γ_k, β_k, and the user-control parameters ρ_k^∞ are also valid here.

For the stability analysis, from the characteristic polynomial of the diagonal-block matrix Λ_k, we define S as

$$S = \frac{2(P-1)\big[2(P-1)(\alpha_k(P-1)+1) - L\big(\alpha_f(P-1)+1\big)\big(2(\alpha_k-\alpha_f)(P-1)-P-1\big)\big]}{(\alpha_f(P-1)+1)\big[(\alpha_k-\alpha_f)(P-1)+P+1\big]^2}. \tag{3.1.38}$$

Based on the definition of the structural dynamics model, we enforce $S \geq 0$. Besides,

considering $\alpha_k \geq \alpha_f \geq \frac{1}{2}$ leads to $-1 < P < 1$ and

$$\begin{aligned}&\left[\left(\frac{1}{2} \leq \alpha_k < \frac{P+5}{4-4P} \;\&\; L \geq -\frac{18(P-1)(\alpha_k(P-1)+1)}{(\alpha_k(P-1)+P+2)(4\alpha_k(P-1)+P+5)}\right)\right.\\ &\text{or}\ \left.\left(\frac{P+5}{4-4P} < \alpha_k \;\&\; L \leq -\frac{18(P-1)(\alpha_k(P-1)+1)}{(\alpha_k(P-1)+P+2)(4\alpha_k(P-1)+P+5)}\right)\right]\\ &\text{or}\ \ (P=1 \;\&\; 2\alpha_k \geq 1)\,.\end{aligned} \tag{3.1.39}$$

The conditions in (3.1.39) show that the method has bounded eigenvalues. Furthermore, there is no condition on choosing L before the spectral radius starts to drift from one (see, Figure 3.2). For the case that $P = \rho_\infty$ which implies $\tau \to 0$, L also can be chosen without any conditions [10]. For the bifurcation interval of spectral radius, we investigate more in remark 10. This completes our analysis.

Remark 9. *The other diagonal blocks have similar structures, and one can readily analyse them. Additionally, the accuracy analysis and the user-control on the spectral behaviour (numerical dissipation) is true (we omit the details for brevity). Therefore, one can set the parameters proposed in* (3.1.25), (3.1.26), *and* (3.1.27) *to gain a desired order of accuracy as well as control the eigenvalues of the amplitude matrix.*

Remark 10. *Following [48], we estimate the eigenvalues of the spatial discretization matrices. Abusing notation, we denote by* $S^2 = \omega^2\tau^2$, *and* ζS *the eigenvalues of the spatial matrices* $\tau^2 \cdot K$ *and* $\tau \cdot C$, *respectively where* ζ *is the damping coefficient and,*

without loss of generality, we set it to 1. Thus, we obtain

$$
\begin{aligned}
\Lambda_k^{11} &= \frac{\alpha_k + S(\alpha_f(\gamma_k + \beta_k S) - \beta_k S)}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{12} &= \frac{\alpha_k - \beta_k S + \alpha_f \gamma_k S}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{13} &= \frac{\alpha_k + \alpha_f \gamma_k S - 2\beta_k(\alpha_f S + 1)}{2(\alpha_k + \alpha_f S(\gamma_k + \beta_k S))}, \\
\Lambda_k^{21} &= -\frac{\gamma_k S^2}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{22} &= \frac{\alpha_k + S(\beta_k \alpha_f S + \gamma_k(-S\alpha_f + \alpha_f - 1))}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{23} &= \frac{2\left(\beta_k \alpha_f S^2 + \alpha_k\right) - \gamma_k\left(\alpha_f S^2 + 2\right)}{2(\alpha_k + \alpha_f S(\gamma_k + \beta_k S))}, \\
\Lambda_k^{31} &= -\frac{S^2}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{32} &= -\frac{S(\alpha_f S + 1)}{\alpha_k + \alpha_f S(\gamma_k + \beta_k S)}, \\
\Lambda_k^{33} &= \frac{2\alpha_k + \alpha_f S(2\gamma_k + 2\beta_k S - S - 2) - 2}{2(\alpha_k + \alpha_f S(\gamma_k + \beta_k S))},
\end{aligned}
\tag{3.1.40}
$$

These expressions are a specific case of (3.1.23)*; therefore, the observations are also valid here. Here, we only analyze the presence of damping when the spectral radius bifurcates from one. Thus, to have positive eigenvalues for the matrix K, we have:*

$$
0 \leq P < 1 \text{ and } \alpha_k < \frac{P+2}{1-P}, \tag{3.1.41}
$$

which is a consequence of definition of α_k using the user-defined parameter ρ_k^∞.

3.1.4.2 Damped system's stability behaviour

The eigenvalues of the stiffness matrix are real and strictly positive as it is a symmetric positive definite matrix, whereas the damping matrix may have complex eigenvalues. The second Dahlquist barrier states that the stable region of a multistep method for a stiff equation shrinks for accuracies higher than two (for more details, see, [48]).

Nevertheless, our method solves k systems identical to the generalized-α method resulting in the amplification matrix of (3.1.24); its blocks are independent of each other, their eigenvalues are independent as well. Therefore, we solve k separate systems, leading to a high-order method with a stability region that is almost unchanged. The

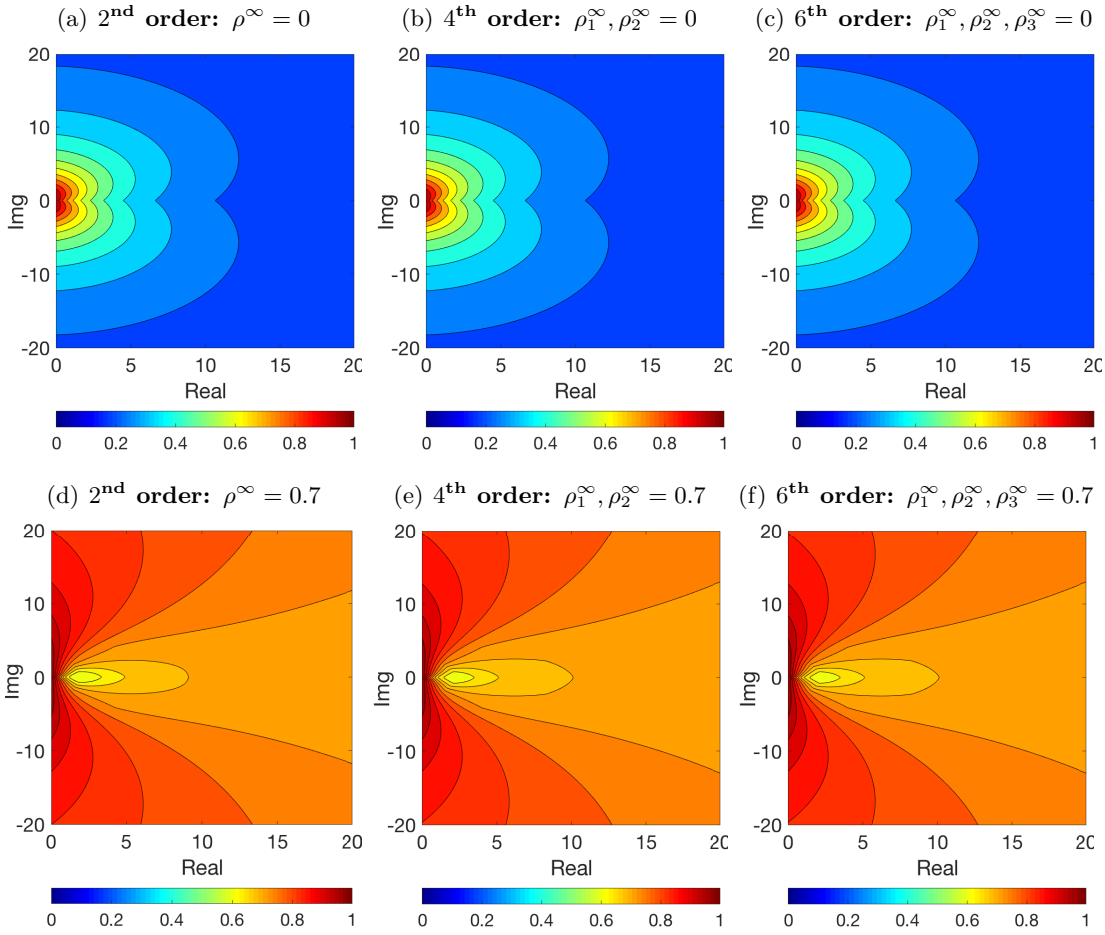


Figure 3.4: Stability region (system's spectral radius) with increasing accuracy-order [10].

accuracy and the stability region are independent; Figure 3.4 shows the boundedness of the system's eigenvalues for damping terms with eigenvalues with real and imaginary parts. The eigenvalue of the stiffness matrix is set equal to the square of the modulus of the damping ones. The largest system's eigenvalue behaves similarly for the second, fourth, and sixth orders of accuracy.

3.2 Explicit higher-order time marching methods for hyperbolic problems

Here, we consider the matrix problem (3.1.32) to introduce an explicit family of high-order time marching schemes. The explicit generalized-α method with second-order

accuracy in time solves (3.1.32) at time-step t_{n+1} by:

$$MA_{n+\alpha_m} + CV_n + KU_n = F_{n+\alpha_f}, \tag{3.2.1a}$$

$$V_{n+1} = V_n + \tau A_n + \tau\gamma[\![A_n]\!], \tag{3.2.1b}$$

$$U_{n+1} = U_n + \tau V_n + \frac{\tau^2}{2}A_n + \tau^2\beta[\![A_n]\!]. \tag{3.2.1c}$$

Similar to the implicit approach, in each time step, we first determine A_{n+1} explicitly using (3.2.1a), and then evaluate V_{n+1} and U_{n+1} from (3.2.1b) and (3.2.1c), respectively. The method (3.2.1) has a truncation error in time of $\mathcal{O}(\tau^3)$ (for further details, see [8, 10]). We extend the accuracy, assuming sufficient temporal regularity of the solution (here, we require defined $\frac{\partial^i u}{\partial t^i}$, $i = 0, \cdots, 5$); thus, we use a Taylor expansion with higher-order terms. Therefore, to derive a fourth-order explicit generalized-α method, we solve

$$\begin{aligned} MA_n^{\alpha_1} + CV_n + KU_n &= F_{n+\alpha_{f1}}, \\ M\mathcal{L}^3(A_n)^{\alpha_2} + C\mathcal{L}^2(A_n) + K\mathcal{L}^1(A_n) &= F^{(3)}_{n+\alpha_{f2}}, \end{aligned} \tag{3.2.2}$$

with updating terms (3.1.5). Using (3.1.35), we also approximate the required initial data. Next, we derive the corresponding coefficients that deliver the desired accuracy and discuss the method's stability.

3.2.1 Order of accuracy in time

We now determine the parameters γ_1 and γ_2 which guarantee the fourth-order accuracy of (3.1.4); thus, we closely follow our logic in Section 3.1.2.1. For this, we first obtain the amplification matrix G with a structure similar to (3.1.23) where Λ_1 and Λ_2 represent the blocks on the diagonal. Then, to produce a truncation error of $\mathcal{O}(\tau^5)$ and consequently obtain a fourth-order accurate scheme in time, we use the following result. Assuming that the solution is sufficiently smooth in time, the method (3.1.4) is fourth-order accurate in the time given (see [12] for further details).:

$$\gamma_1 = \frac{1}{2} - \alpha_{f1} + \alpha_1, \qquad \gamma_2 = \frac{1}{2} - \alpha_{f2} + \alpha_2. \tag{3.2.3}$$

3.2.2 Stability analysis and CFL condition

Due to the conditional stability of the explicit methods, we derive the corresponding stability conditions. For this, we bound the spectral radius of the amplification matrix by one. Therefore, we calculate the eigenvalues of the amplification matrix G of the

method introduced in (3.2.2) for the case $\Theta := \tau^2\lambda \to 0$ where λ is the largest eigenvalue corresponding to the spatial discretization. Thus, in this case, G's diagonal blocks read:

$$\Lambda_1 = \begin{bmatrix} 1 & 1 & \frac{1}{2} - \frac{\beta_1}{\alpha_1} \\ 0 & 1 & 1 - \frac{\gamma_1}{\alpha_1} \\ 0 & 0 & 1 - \frac{1}{\alpha_1} \end{bmatrix}, \; \Lambda_2 \; = \begin{bmatrix} 1 & 1 & \frac{1}{2} - \frac{\beta_2}{\alpha_2} \\ 0 & 1 & 1 - \frac{\gamma_2}{\alpha_2} \\ 0 & 0 & 1 - \frac{1}{\alpha_2} \end{bmatrix}. \tag{3.2.4}$$

Then, G's eigenvalues are:

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1, \qquad \lambda_5 = 1 - \frac{1}{\alpha_1}, \qquad \lambda_6 = 1 - \frac{1}{\alpha_2}. \tag{3.2.5}$$

The boundedness of λ_5 and λ_6 in (3.2.5) implies that $\alpha_1 \geq \frac{1}{2}$ and $\alpha_2 \geq \frac{1}{2}$.

Next, we introduce the CFL condition for our scheme. In the analysis of the implicit generalized-α methods of second- and higher-order accuracy [10, 26], we analyze the discrete system's eigenvalue distribution in the limit $\Theta \to \infty$ and set the method's free parameters such that all eigenvalues are equal to a real constant $\rho_\infty \in [0, 1]$; this parameter controls the numerical dissipation. In our explicit method, we need to find the method's conditional stability region. Thus, we exploit the amplification matrix's upper-triangular structure to explicitly compute the eigenvalues of G, which is equivalent to finding the eigenvalues of each diagonal block (see, [12]). Therefore, the characteristic polynomials for the diagonal blocks are:

$$\begin{aligned} \Lambda_1 : \quad & 2\alpha_1\lambda_1^3 + \lambda_1^2\left(\Theta(2\beta_1 + 2\gamma_1 + 1) - 6\alpha_1 + 2\right) \\ & \quad + \lambda_1\left(\Theta(-4\beta_1 - 2\gamma_1 + 1) + 6\alpha_1 - 4\right) + 2\beta_1\Theta - 2\alpha_1 + 2 = 0, \\ \Lambda_2 : \quad & 2\alpha_2\lambda_2^3 + \lambda_2^2\left(2\Theta\beta_2 - 6\alpha_2 + 2\right) + \lambda_2\left(\Theta(-4\beta_2 + 2\gamma_2 + 1) + 6\alpha_2 - 4\right) \\ & \quad + \Theta(2\beta_2 + 1 - 2\gamma_2) - 2\alpha_2 + 2 = 0, \end{aligned} \tag{3.2.6}$$

where two roots of each characteristic polynomial in (3.2.6) are the principal roots while the third is spurious (unphysical).; thus, we require the two principal roots to be complex conjugates except in the high-frequency regions. This requirement maximizes high-frequency dissipation while setting two eigenvalues to one in the low-frequency range to improve its approximation accuracy. As a consequence, we set β_1 and β_2 such that the complex parts of the principal roots of the blocks Λ_1 and Λ_2, respectively, vanish in the high-frequency regions. This setting changes the largest eigenvalue from one to a user-defined value, which results in an approach similar to the implicit generalized-α methods. Additionally, we define the limit in which this bifurcation happens at block

i, $i = 1, 2$, by Ω_{bi}. We define the critical stability limit for our explicit method as Ω_{si} and show the stability region of block i.

We find the parameter values using the characteristic equation corresponding to the diagonal block i of the amplification matrix as:

$$\sum_{j=0}^{3} \left(\tilde{a}_i^j + \tilde{y}_i^j \Theta^2 \right) \lambda_i^{(3-j)} = 0, \tag{3.2.7}$$

where $\tilde{a}_i$ and $\tilde{y}_i$ are functions of the parameters γ_i and β_i. Each block i has three eigenvalues; we set two of them to ρ_{bi} and one becomes ρ_{si}. Therefore, we rewrite the characteristic polynomial (4.2.21) as

$$\left(\lambda_i^2 + \rho_{bi}^2 + 2\lambda_i \rho_{bi} \right) (\lambda_i + \rho_{si}) = 0. \tag{3.2.8}$$

Next, to have all three roots with real values at the bifurcation limit, we equate (3.2.8) to (3.2.6) to obtain:

$$\begin{aligned} \Omega_{b1} &= 2 - 2\rho_{b1} - \rho_{s1} + \rho_{s1}\rho_{b1}^2, \\ \Omega_{b2} &= 2 + 2\rho_{b2} + \rho_{s2} - \rho_{s2}\rho_{b2}^2, \\ \alpha_1 &= \frac{2 - (1 + \rho_{b1})\rho_{s1}}{(-1 + \rho_{b1})(-1 + \rho_{s1})} \\ \alpha_2 &= \frac{2 + (1 - \rho_{b2})\rho_{s2}}{(1 + \rho_{b2})(1 + \rho_{s2})}. \end{aligned} \tag{3.2.9}$$

Figure 3.5 shows how the bifurcation and stability regions change as the user-defined parameter changes.

Remark 11. *Herein, we constrain $\rho_{si} \leq \rho_{bi}$ which results in $\Omega_{bi} \leq \Omega_{si}$. We can maximize the bifurcation region Ω_{bi} by setting $\rho_{si} = \rho_{bi}$ and consequently, obtain a one-parameter family of algorithms, Figure 3.5 shows this numerically.*

We set the complex part of the eigenvalues equal to zero in the high-frequency regions; thus, we define the parameters β_i as

$$\begin{aligned} \beta_1 &= \frac{(1 + \rho_{b1})(-1 + \rho_{b1}\rho_{s1})^2}{(-1 + \rho_{b1})^2(-1 + \rho_{s1})(-2 + \rho_{s1} + \rho_{b1}\rho_{s1})}, \\ \beta_2 &= \frac{-5 - 3\rho_{b2} - 4\rho_{s2} + 2\rho_{b2}\rho_{s2} + 2\rho_{b2}^2\rho_{s2} - \rho_{s2}^2 + \rho_{b2}\rho_{s2}^2}{(1 + \rho_{b2})^2(-2 - 3\rho_{s2} + \rho_{b2}\rho_{s2} - \rho_{s2}^2 + \rho_{b2}\rho_{s2}^2)}. \end{aligned} \tag{3.2.10}$$

Finally, we find the critical values at which each block's spectral radius becomes larger than one. For this, we set $\Theta = \Omega_{si}$ and $\lambda_i = 1$ in (4.2.21) and solve the resulting

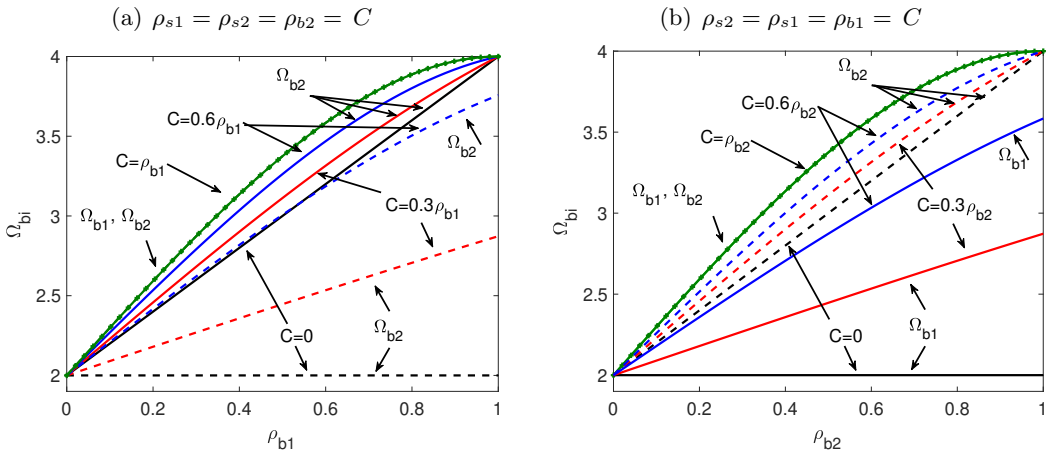


Figure 3.5: Effect of user-defined parameters on the bifurcation regions Ω_{b1} and Ω_{b2} [12].

equation. Thus, we introduce the critical values Ω_{si} as

$$\begin{aligned}\Omega_{s1} &= \frac{4(1-\rho_{b1})(2-\rho_{b1}\rho_{s1}-\rho_{s1})(3+\rho_{b1}-\rho_{s1}-3\rho_{b1}\rho_{s1})}{2(5-\rho_{b1}^2)+(5-13\rho_{b1}-\rho_{b1}^2-\rho_{b1}^3)\rho_{s1}-(1+\rho_{b1})^3\rho_{s1}^2},\\ \Omega_{s2} &= \frac{4(1+\rho_{b2})(2-\rho_{b2}\rho_{s2}+\rho_{s2})(3-\rho_{b2}+\rho_{s2}-3\rho_{b2}\rho_{s2})}{2(5-\rho_{b2}^2)+(5-13\rho_{b2}-\rho_{b2}^2+\rho_{b2}^3)\rho_{s2}-(1-\rho_{b2})^3\rho_{s2}^2}.\end{aligned} \tag{3.2.11}$$

Remark 12. *Following [7, 57], α_{f1} and α_{f2} are free parameters with respect to spectral radius. Therefore, we set $\alpha_{f1} = 1$ to deliver fourth-order accuracy. Accordingly, to optimally combine low- and high-frequency dissipation, we set $\alpha_{f2} = 0$.*

Therefore, setting $0 \leq \rho_{b1}, \rho_{b2} \leq 1$ allows us to control the system's spectral radius ρ and, consequently, the high-frequency numerical damping. Figure 3.6 shows how the user controls ρ and the stability region Ω_s; setting $\rho_b = \rho_s = 0.99$ leads to the largest stability region $\Omega_s = 4(\theta\tau)^2 = 4\Theta$, equivalent to the stability region of the second-order central difference method [48].

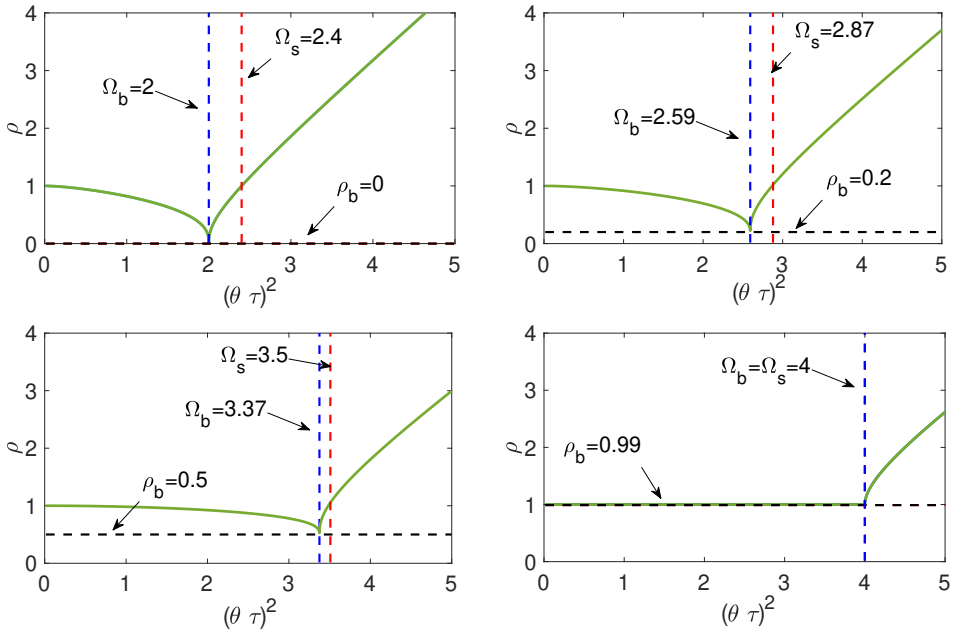


Figure 3.6: Spectral radius ρ for different ρ_b values, where $\rho_{b1} = \rho_{b2} = \rho_{s1} = \rho_{s2}$ [12].

3.2.3 Higher-order explicit method: $2k^{th}$-order accuracy in time

This section extends our approach to deliver $2k^{th}$-order accurate methods with $k \geq 2$. That is, we introduce our $2k^{th}$-order explicit generalized-α method as:

$$\begin{aligned} MA_n^{\alpha_1} + CV_n + KU_n &= F_{n+\alpha_{f1}}, \\ M\mathcal{L}^{3j-3}(A_n)^{\alpha_j} + C\mathcal{L}^{3j-4}(A_n) + K\mathcal{L}^{3j-5}(A_n) &= F^{(3j-3)}_{n+\alpha_{fj}}, , \qquad j = 2, \cdots, k-1, \\ M\mathcal{L}^{3k-3}(A_n)^{\alpha_k} + C\mathcal{L}^{3k-4}(A_n) + K\mathcal{L}^{3k-5}(A_n) &= F^{(3k-3)}_{n+\alpha_{fk}}, \end{aligned} \tag{3.2.12}$$

and updating the system using (3.1.20) with the definitions given in (3.1.21). Similarly, we establish higher-order schemes proposed in (3.2.12) and (3.1.20) by using high-order Taylor series. Next, for the analysis, we substitute (3.1.20) into (3.2.12) and find a matrix system similar to (3.1.36) where the amplification matrix of the $2k^{th}$-order accurate scheme is:

$$G_k = L^{-1}R = \begin{bmatrix} \Lambda_1 & \Xi_{12} & \cdots & \cdots & \Xi_{1k} \\ 0 & \Lambda_2 & \Xi_{23} & \cdots & \Xi_{2k} \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \Lambda_{k-1} & \Xi_{k-1} \\ 0 & 0 & \cdots & 0 & \Lambda_k \end{bmatrix}, \tag{3.2.13}$$

with

$$\begin{aligned}
\Lambda_j &= \frac{1}{\alpha_j}\begin{bmatrix} \alpha_j-\beta_j\Theta & \alpha_j-\beta_j\Theta & \frac{1}{2}\left(\alpha_j-\beta_j\left(\Theta+2\right)\right) \\ -\gamma_j\Theta & \alpha_j-\gamma_j\Theta & \alpha_j-\frac{\gamma_j}{2}\left(\Theta+2\right) \\ -\Theta & -\Theta & \alpha_j-1-\frac{\Theta}{2} \end{bmatrix}, \qquad j=1,\cdots,k-1 \\
\Lambda_k &= \frac{1}{\alpha_k}\begin{bmatrix} \alpha_k-\beta_k\Theta & \alpha_k & \frac{\alpha_k}{2}-\beta_k \\ -\gamma_k\Theta & \alpha_k & \alpha_k-\gamma_k \\ -\Theta & 0 & \alpha_k-1 \end{bmatrix}.
\end{aligned} \tag{3.2.14}$$

Remark 13. *The solution $u(\cdot\,,t)$ nees to be sufficiently regularity such that $\dfrac{\partial^i u}{\partial t^i}$, $i=0,\cdots,3k-1$ are defined for us to achieve $2k^{th}$-order of accuracy using* (3.2.12)*, Otherwise, one needs to estimate the higher-order terms used in the Taylor's expansions using the combination of lower-order terms in different time steps. This alternative approach is similar to multi-step methods such as BDF schemes (see, e.g., [93]).*

Remark 14. *The amplification matrix* (3.2.13) *is an upper-triangular block matrix; thus, we neglect the non-diagonal block contributions in the eigenvalue analysis (see the sketch in Figure 3.3).*

Theorem 1. *Assuming $u_h(t)$ and $\dot{u}_h(t)$ have sufficient regularity in time, our explicit scheme introduced in* (3.2.12) *with updating terms* (3.1.20) *provides $2k^{th}$-order accurate approximation in the temporal domain when*

$$\begin{aligned}
\gamma_j &= \frac{1}{2}-\alpha_{fj}+\alpha_j, \qquad j=1,\cdots,k-1, \\
\gamma_k &= \frac{1}{2}-\alpha_{fk}+\alpha_k.
\end{aligned} \tag{3.2.15}$$

Proof. The amplification matrix (3.2.13) is an upper-diagonal block matrix; each block is a 3×3 matrix. The last two diagonal blocks Λ_{k-1} and Λ_k have similar entries to those of the amplification matrix of the fourth-order method. The other diagonal blocks are analogous to Λ_{k-1}. Hence, separately for each block, we determine the relevant terms for (3.1.11) and consider the higher-order terms to obtain second-order accuracy. Consequently, after solving the whole system and adding the higher-order terms to the unknowns u_h and v_h, we have a truncation error of $\mathcal{O}(\tau^{2k+1})$. A similar path we used in Section 3.1.3. □

3.2.3.1 CFL condition and dissipation control

We analyze the system's spectral behaviour, the bifurcation regions, and its CFL conditions. For this, we study the amplification matrix's eigenvalues (3.2.13). Accordingly,

we calculate each diagonal block's eigenvalues. The first $k-1$ blocks have identical structures; hence, following Section 3.2.2, we propose these algorithmic parameters as

$$\boxed{\begin{aligned}
\Omega_{bj} &= 2 - 2\rho_{bj} - \rho_{sj} + \rho_{sj}\rho_{bj}^2, \qquad j = 1, \cdots, k-1, \\
\alpha_j &= \frac{2 - (1+\rho_{bj})\rho_{sj}}{(-1+\rho_{bj})(-1+\rho_{sj})}, \\
\beta_j &= \frac{(1+\rho_{bj})(-1+\rho_{bj}\rho_{sj})^2}{(-1+\rho_{bj})^2(-1+\rho_{sj})(-2+\rho_{sj}+\rho_{bj}\rho_{sj})}, \\
\Omega_{bk} &= 2 + 2\rho_{bk} + \rho_{sk} - \rho_{sk}\rho_{bk}^2, \\
\alpha_k &= \frac{2 + (1-\rho_{bk})\rho_{sk}}{(1+\rho_{bk})(1+\rho_{sk})}, \\
\beta_k &= \frac{-5 - 3\rho_{bk} - 4\rho_{sk} + 2\rho_{bk}\rho_{sk} + 2\rho_{bk}^2\rho_{sk} - \rho_{sk}^2 + \rho_{bk}\rho_{sk}^2}{(1+\rho_{bk})^2(-2 - 3\rho_{sk} + \rho_{bk}\rho_{sk} - \rho_{sk}^2 + \rho_{bk}\rho_{sk}^2)},
\end{aligned}} \tag{3.2.16}$$

and the critical values Ω_{sj} are

$$\begin{aligned}
\Omega_{sj} &= \frac{4(1-\rho_{bj})(2-\rho_{bj}\rho_{sj}-\rho_{sj})(3+\rho_{bj}-\rho_{sj}-3\rho_{bj}\rho_{sj})}{2(5-\rho_{bj}^2) + (5-13\rho_{bj}-\rho_{bj}^2-\rho_{bj}^3)\rho_{sj} - (1+\rho_{bj})^3\rho_{sj}^2}, \quad j = 1, \cdots, k-1, \\
\Omega_{sk} &= \frac{4(1+\rho_{bk})(2-\rho_{bk}\rho_{sk}+\rho_{sk})(3-\rho_{bk}+\rho_{sk}-3\rho_{bk}\rho_{sk})}{2(5-\rho_{bk}^2) + (5-13\rho_{bk}-\rho_{bk}^2+\rho_{bk}^3)\rho_{sk} - (1-\rho_{bk})^3\rho_{sk}^2}.
\end{aligned} \tag{3.2.17}$$

Remark 15. *We constrain $\rho_{sm} \leq \rho_{bm}$, $m = 1, \cdots, k$ and maximize the bifurcation regions Ω_{bm} by setting $\rho_{sm} = \rho_{bm} = \rho$ with $0 \leq \rho < 1$ as a user-defined parameter; our method is a one-parameter family of time-marching algorithms. Additionally, the stability region is independent of the accuracy order; we obtain higher-order accuracy without affecting any features of the second-order algorithm (i.e., preserve stability regions, bifurcation limit, and dissipation control). Similarly, α_{fj} and α_{fk} are free parameters; thus, we set $\alpha_{fj} = 1$, $j = 1, \cdots, k-1$, and $\alpha_{fk} = 0$.*

3.3 Implicit higher-order time marching methods for parabolic problems

3.3.1 Model problem

Here, we introduce our method for approximating first-order time derivatives. For this, we consider a parabolic, linear initial boundary-value problem:

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} - \nabla \cdot (\kappa \nabla u(x,t)) &= f(x,t), && (x,t) \in \Omega \times [0\,,\,T], \\ u(x,t) &= u_D, && x \in \partial\Omega \times [0\,,\,T], \\ u(x,0) &= u_0, && x \in \Omega, \end{aligned} \tag{3.3.1}$$

where $\kappa \in L^\infty(\Omega)$ is the diffusivity coefficient, the source function f, the initial data u_0, and the Dirichlet boundary condition u_D are given.

For the spatial discretization, we follow the steps discussed in Section 3.1.4. Thus, the matrix form of the discrete problem (3.3.1) becomes:

$$M\dot{U} + KU = F, \tag{3.3.2}$$

where M and K are the mass and stiffness matrices, U is the vector of the unknowns, and F is the source vector. The initial condition is

$$U(0) = U_0, \tag{3.3.3}$$

where U_0 is the given vector of initial condition $u_{0,h}$.

Remark 16. *Similar to our method for hyperbolic problems, we propose a high-order generalized-α independent of the spatial discretization. Therefore, one can apply our construction to any time-dependent semi-discrete problem. Besides, the problem* (3.3.1) *can be easily modified to consider heterogeneous diffusivity and non-homogeneous boundary conditions (for more details, see [56]).*

3.3.2 Generalized-α for parabolic problems

Partitioning of the time interval $[0, T]$, the generalized-α method for (3.3.2) at time-step $n+1$ reads [60]:

$$\begin{aligned} MV_{n+\alpha_m} + KU_{n+\alpha_f} &= F_{n+\alpha_f}, \\ U_{n+1} &= U_n + \tau V_n + \tau\gamma[\![V_n]\!], \\ V_0 &= M^{-1}(F_0 - KU_0), \end{aligned} \tag{3.3.4}$$

According to (3.3.4), the method requires a two-step computation; the first one solves an implicit system to find $[\![V_n]\!]$, the second one deploys the second equation in (3.3.4)

to update U_{n+1} explicitly. We guarantee the scheme's second-order accuracy in time by setting $\gamma = \frac{1}{2} + \alpha_m - \alpha_f$; we also control the numerical dissipation using the following parameter definition [60]:

$$\alpha_m = \frac{1}{2}\Big(\frac{3-\rho_\infty}{1+\rho_\infty}\Big), \qquad \alpha_f = \frac{1}{1+\rho_\infty}. \tag{3.3.5}$$

where, $\rho_\infty \in [0, 1]$ is a user-control parameter.

3.3.3 Third-order generalized-α method

Similar to our approach for developing high-order schemes for hyperbolic problems, we use a Taylor expansion to extend the second-order method (3.3.4) and derive higher-order representations; see also our detailed discussions in [8, 10, 12]. Thus, for example, we derive a third-order generalized-α method for solving the semi-discrete equation (3.3.2) which has a sufficiently smooth solution and forcing on the considered time interval. For this, taking two temporal derivatives from the first equation of (3.3.4), we $M\mathcal{L}^1(A_n)^{\alpha_2} + KA_n^{\alpha_f} = F_{n+\alpha_f}^{(2)}$. Then, we propose:

$$\begin{aligned} MV_n^{\alpha_1} &= -KU_{n+1} + F_{n+1}, \\ M\mathcal{L}^1(A_n)^{\alpha_2} &= -KA_n^{\alpha_f} + F_{n+\alpha_f}^{(2)}, \end{aligned} \tag{3.3.6}$$

with updating conditions

$$\begin{aligned} U_{n+1} &= U_n + \tau V_n + \frac{\tau^2}{2}A_n + \frac{\tau^3}{6}\mathcal{L}^1(A_n) + \gamma_1\tau Q_n, \\ A_{n+1} &= A_n + \tau\mathcal{L}^1(A_n) + \tau\gamma_2[\![\mathcal{L}^1(A_n)]\!], \end{aligned} \tag{3.3.7}$$

where

$$\begin{aligned} Q_n &= V_{n+1} - V_n - \tau A_n - \frac{\tau^2}{2}\mathcal{L}^1(A_n), \\ V_n^{\alpha_1} &= V_n + \tau A_n + \frac{\tau^2}{2}\mathcal{L}^1(A_n) + \alpha_1 Q_n. \end{aligned} \tag{3.3.8}$$

The other requires data at the initial state are determined using the given initial condition U_0 as:

$$\begin{aligned} V_0 &= M^{-1}(F_0 - KU_0), \\ A_0 &= M^{-1}(\mathcal{L}^1(F_0) - KV_0), \\ \mathcal{L}^1(A_0) &= M^{-1}(\mathcal{L}^2(F_0) - KA_0). \end{aligned} \tag{3.3.9}$$

Next, we define the parameters to assure stability and the third-order of accuracy.

3.3.3.1 Accuracy in time

Now, we obtain the parameters γ_1 and γ_2 such that equations (3.3.6)-(3.3.7) deliver third-order accuracy in the temporal domain, which renders the following result:

Theorem 2. *Considering a problem with a sufficiently regular solution, the method in* (3.3.6) *with the update* (3.3.7) *is third-order accurate in the time given*

$$\gamma_1 = \alpha_1 - \frac{1}{2}, \qquad \gamma_2 = \frac{1}{2} - \alpha_f + \alpha_2. \tag{3.3.10}$$

Proof. Substituting (3.3.7) into (3.3.6), we obtain an equation system for each time step similar to (3.1.14) as:

$$A\mathbf{U_{n+1}} = B\mathbf{U_n} + \mathbf{F_{n+\alpha_f}}, \tag{3.3.11}$$

For simplicity, we assume the matrix problem (3.3.11) has one spatial degree of freedom. Then, letting $\mathbf{U}_n = \left[U_n,\ \tau V_n,\ \tau^2 A_n,\ \tau^3 \mathcal{L}^1(A_n)\right]^T$, we have:

$$A = \begin{bmatrix} 1 & -\gamma_1 & 0 & 0 \\ \tau\lambda & \alpha_1 & 0 & 0 \\ 0 & 0 & 1 & -\gamma_2 \\ 0 & 0 & \tau\alpha_f\lambda & \alpha_2 \end{bmatrix},\ B = \begin{bmatrix} 1 & 1-\gamma_1 & \frac{1}{2}-\gamma_1 & \frac{1}{6}-\frac{1}{2}\gamma_1 \\ 0 & \alpha_1 - 1 & \alpha_1 - 1 & \frac{1}{2}(\alpha_1 - 1) \\ 0 & 0 & 1 & 1-\gamma_2 \\ 0 & 0 & -\tau(1-\alpha_f)\lambda & \alpha_2 - 1 \end{bmatrix}. \tag{3.3.12}$$

Remark 17. *In our analysis, the behaviour of the time-marching methods is independent of spatial discretization. This leads to the assumption of considering one degree of freedom for the spatial discretization for brevity. Whereas, in chapter 4 that we introduce our variational splitting techniques, we require to also include the eigenvalues of the matrices obtained from the spatial discretization. Therefore, we analyse the systems using an alternative approach (see, e.g., 4.2.2).*

Without loss of generality, we set $\mathbf{F_{n+\alpha_f}} = \mathbf{0}$. Thus, the resulting amplification matrix G is similar to the one in (3.1.10) with an upper-block diagonal matrix:

$$G = \begin{bmatrix} \Lambda_1 & \Xi \\ 0 & \Lambda_2 \end{bmatrix}, \tag{3.3.13}$$

with

$$\Lambda_1 = \theta_1 \begin{bmatrix} \alpha_1 & \alpha_1 - \gamma_1 \\ -\theta & \alpha_1 + (\gamma_1 - 1)\theta - 1 \end{bmatrix}, \tag{3.3.14}$$

$$\Lambda_2 = \theta_2 \begin{bmatrix} \alpha_2 + (\alpha_f - 1)\gamma_2\theta & \alpha_2 - \gamma_2 \\ -\theta & \alpha_2 + \alpha_f(\gamma_2 - 1)\theta - 1 \end{bmatrix}, \tag{3.3.15}$$

where $\theta_1 = (\alpha_1 + \gamma_1\theta)^{-1}$ and $\theta_2 = (\alpha_2 + \alpha_f\gamma_2\theta)^{-1}$. Additionally, we define $\theta := \tau\lambda_\theta$ with λ_θ being the eigenvalues of $M^{-1}K$ related to the spatial discretization. Then, for the 4×4 matrix G, we can write:

$$\begin{aligned} 0 = U^{n+1} - (\text{tr } G)U^n + \frac{1}{2}\left((\text{tr } G)^2 - \text{tr } (G^2)\right)U^{n-1} \\ - \frac{1}{6}\left((\text{tr } G)^3 - 3\text{tr } (G^2)(\text{tr } G) + 2\text{tr } (G^3)\right)U^{n-2} + \det(G)U^{n-3}, \end{aligned} \tag{3.3.16}$$

The identity (3.3.16) is a consequence of Cayley-Hamilton Theorem for a 4×4 matrix. We detail these computations when we analyze a general $k \times k$ matrix in the next section. Later, we substitute the Taylor expansions of U^{n+1}, U^{n-1}, U^{n-2} in time with the truncation error of $\mathcal{O}(\tau^4)$ into (3.3.16). Then, one can verify that these parameter definitions (3.3.10) cancel the low-order terms and the method delivers the truncation error of fourth-order and therefore, third-order accuracy in the temporal domain [9]. □

Remark 18. *The standard generalized-α method [60] delivers second-order accuracy and dissipation control using two equations; similarly, we obtain a third-order rather than fourth-order to maintain the unconditional stability and dissipation control. For example, we get fourth-order accuracy by setting $\gamma_2 = 0$, which results in an explicit-implicit method with a CFL condition. Therefore, we omit the details for other possible choices of γ_i and only consider* (3.3.10) *that leads to unconditional A-stability.*

Remark 19 (Alternative proof for accuracy order)**.** *As with the hyperbolic schemes, due to G's structure, we can only study the diagonal blocks' behaviour in our analysis. Therefore, we require that the higher-order unknowns A_n, $\mathcal{L}^1(A_n)$, associated with the lower diagonal-block Λ_2 be second-order accurate (see, Section 3.1.2.1 and [9, 10]):*

$$A_{n+1} - tr\,(\Lambda_2)\,A_n - \det(\Lambda_2)\,A_{n-1} = 0. \tag{3.3.17}$$

Then, further substitution of Taylor expansions of A_{n+1} and A_{n-1} in time as:

$$\begin{aligned} A_{n+1} &= A_n + \tau\mathcal{L}^1(A_n) + \frac{\tau^2}{2}\mathcal{L}^2(A_n) + \mathcal{O}(\tau^3), \\ A_{n-1} &= A_n - \tau\mathcal{L}^1(A_n) + \frac{\tau^2}{2}\mathcal{L}^2(A_n) + \mathcal{O}(\tau^3), \end{aligned} \tag{3.3.18}$$

requires $\gamma_2 = \frac{1}{2} - \alpha_f + \alpha_2$ to guarantee the second-order accuracy in time of A_{n+1}. Next, according to (3.3.17) *and deploying the upper diagonal-block Λ_1, we write:*

$$U_{n+1} - tr\,(\Lambda_1)\,U_n + \det(\Lambda_1)\,U_{n-1} = 0. \tag{3.3.19}$$

We obtain the third-order accuracy in time using Taylor expansions with truncation error of $\mathcal{O}(\tau^4)$ for U_{n+1} and U_{n-1}, which allows to get

$$\begin{aligned} U_{n+1} &= U_n + \tau V_n + \underbrace{\frac{\tau^2}{2}A_n + \frac{\tau^3}{6}\mathcal{L}^1(A_n)}_{\mathcal{R}} + \mathcal{O}(\tau^4), \\ U_{n-1} &= U_n - \tau V_n + \underbrace{\frac{\tau^2}{2}A_n - \frac{\tau^3}{6}\mathcal{L}^1(A_n)}_{\mathcal{R}} + \mathcal{O}(\tau^4), \end{aligned} \tag{3.3.20}$$

Here, the Λ_2 equations show that $\mathcal{R}$ has second-order accuracy and define a residual term. Thus, following a similar approach, we verify that the remaining terms have second-order accuracy in time by setting $\gamma_1 = \alpha_1 - \frac{1}{2}$. Then, we add the residuals to the second-order accurate solution to have the truncation error of $\mathcal{O}(\tau^4)$ and consequently, a third-order accurate scheme in time, which completes the proof [9].

3.3.3.2 Stability analysis and eigenvalue control

Following the methodology of Section 3.1.2.2, we bound the spectral radius of the amplification G; thus, we first calculate the eigenvalues of G as:

$$\begin{aligned} 0 = \det\left(G - \tilde{\lambda}I\right) &= \det\begin{bmatrix} \Lambda_1 - \tilde{\lambda}I & H \\ 0 & \Lambda_2 - \tilde{\lambda}I \end{bmatrix} \\ &= \det\left(\Lambda_1 - \tilde{\lambda}I\right) \cdot \det\left(\Lambda_2 - \tilde{\lambda}I\right), \end{aligned} \tag{3.3.21}$$

with I and $\tilde{\lambda}$ denoting the corresponding identity matrix and eigenvalues of the matrix, respectively. Thus, we obtain a solution for (3.3.21) by solving two uncoupled problems $\det\left(\Lambda_i - \tilde{\lambda}I\right) = 0$, $i = 1, 2$. For this, we have [53]:

$$\det\left(\Lambda_i - \tilde{\lambda}I\right) = \det\left(\Lambda_i^{11} - \tilde{\lambda}I\right) \cdot \det\left(\Lambda_i^{22} - \tilde{\lambda}I - \Lambda_i^{21}\left(\Lambda_i^{11} - \tilde{\lambda}I\right)^{-1}\Lambda_i^{12}\right), \tag{3.3.22}$$

where Λ_i^{jk} is the jk component of Λ_i. Therefore, substituting (3.3.15) into (3.3.22), we consider $\det\left(\Lambda_2^{11} - \tilde{\lambda}I\right) = 0$ which leads to the following bound on $\tilde{\lambda}$:

$$-1 \leq (\alpha_2 + (\alpha_f - 1)\gamma_2\theta)\,\theta_2 \leq 1 \implies 0 \leq \gamma_2\theta\theta_2 \leq 2. \tag{3.3.23}$$

The left inequality is already satisfied since all parameters are non-negative, and the matrices M and K obtained after the spatial discretization are positive definite. We rewrite the right-hand side of the inequality as:

$$\gamma_2\theta(1 - 2\alpha_f) \leq 2\alpha_2. \tag{3.3.24}$$

To satisfy (3.3.24) unconditionally, one requires $\frac{1}{2} \leq \alpha_f$. Next, we solve

$$\det\left(\Lambda_2^{22} - \tilde{\lambda} I - \Lambda_2^{21}\left(\Lambda_2^{11} - \tilde{\lambda} I\right)^{-1} \Lambda_2^{12}\right) = 0$$

in (3.3.22) using:

$$\left(1 - \gamma_2\theta\theta_2\tilde{\lambda}\right) \cdot \left(1 - (1 + \alpha_f\gamma_2\theta)\theta_2\tilde{\lambda}\right) + (1 - \theta\theta_2) \cdot (1 - (\gamma_2 + \alpha_f\gamma_2\theta)\theta_2) = 0. \quad (3.3.25)$$

Omitting the details, we bound the spectral radius by one; it is sufficient to impose [9, 26]:

$$\alpha_2 \geq \alpha_f \geq \frac{1}{2}. \quad (3.3.26)$$

To solve (3.3.22) for $i = 1$, one can follow the same steps. Thus, we solve

$$\det\left(\Lambda_1^{11} - \tilde{\lambda} I\right) = 0$$

and bound the spectral radius as:

$$-1 \leq \alpha_1\theta_1 \leq 1 \implies \gamma_1\theta \leq 2\left(\alpha_1 + \gamma_1\theta\right). \quad (3.3.27)$$

Furthermore, substituting

$$\det\left(\Lambda_1^{22} - \tilde{\lambda} I - \Lambda_1^{21}\left(\Lambda_1^{11} - \tilde{\lambda} I\right)^{-1} \Lambda_1^{12}\right) = 0$$

in (3.3.22) implies that $\alpha_1 \geq 1$. Again, we omit the details for brevity. We calculate the eigenvalues for the case $\theta \to \infty$ equal to ρ_1^∞ to control the eigenvalues in high frequency regions. Accordingly, the eigenvalues of the amplification matrix read:

$$\lambda_1 = 0, \quad \lambda_2 = \frac{\gamma_1 - 1}{\gamma_1}, \quad \lambda_3 = \frac{\alpha_f - 1}{\alpha_f}, \quad \lambda_4 = \frac{\gamma_2 - 1}{\gamma_2}. \quad (3.3.28)$$

Providing control on the numerical dissipation, we closely follow the analysis in [7, 26, 60], we set $\lambda_2 = \rho_1^\infty$ and $\lambda_3 = \lambda_4 = \rho_2^\infty$ and find corresponding expressions for α_1, α_2, α_f as:

$$\begin{aligned} \alpha_1 &= \frac{1}{2}\left(\frac{3 + \rho_1^\infty}{1 + \rho_1^\infty}\right), \\ \alpha_2 &= \frac{1}{2}\left(\frac{3 - \rho_2^\infty}{1 + \rho_2^\infty}\right), \\ \alpha_f &= \frac{1}{1 + \rho_2^\infty}. \end{aligned} \quad (3.3.29)$$

Therefore, setting $0 \leq \rho_1^\infty, \rho_2^\infty \leq 1$, one controls the eigenvalues of the amplification matrix and the high-frequency damping. Figure 3.7 shows the behaviour of these eigenvalues; for large θ, the eigenvalues $\lambda_{1,2}$ of the first block of the amplification matrix approach 0 and ρ_1^∞ and eigenvalues of the second block $\lambda_{3,4}$ converge to ρ_2^∞.

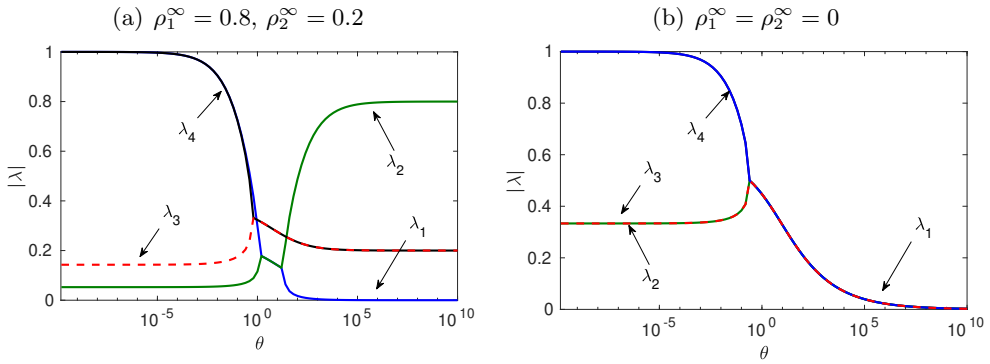


Figure 3.7: Amplification matrix eigenvalues, G, (3.3.13).

3.3.4 Higher-order accuracy in time

This section proposes a method that attains higher-order accuracy in the temporal domain to solve the parabolic problem. In general, to solve the semi-discrete matrix problem (3.3.2), we introduce a high-order generalized-α method by solving $k \geq 2$ equations as [9]:

$$\begin{aligned} MV_n^{\alpha_1} &= -KU_{n+1} + F_{n+1}, \\ M\mathcal{L}^{2j-3}(A_n)^{\alpha_j} &= -K\mathcal{L}^{2j-4}(A_{n+1}) + \mathcal{L}^{2j-2}(F_{n+1}), \qquad j = 2, \cdots, k-1, \\ M\mathcal{L}^{2k-3}(A_n)^{\alpha_k} &= -K\mathcal{L}^{2k-4}(A_n)^{\alpha_f} + \mathcal{L}^{2k-2}(F_n)^{\alpha_f}, \end{aligned} \tag{3.3.30}$$

to obtain $(\frac{3}{2}k)^{th}$ order of accuracy for even k and $(\frac{3}{2}k+\frac{1}{2})^{th}$ for odd k. Next, we update the system explicitly using the following expressions:

$$\begin{aligned} U_{n+1} &= U_n + \tau V_n + \frac{\tau^2}{2}A_n + \frac{\tau^3}{6}\mathcal{L}^1(A_n) + \cdots + \frac{\tau^{2k-1}}{(2k-1)!}\mathcal{L}^{2k-3}(A_n) + \tau\gamma_1 Q_{n,1}, \\ \mathcal{L}^{2j-4}(A_{n+1}) &= \mathcal{L}^{2j-4}(A_n) + \cdots + \frac{\tau^{2k-2j+1}}{(2k-2j+1)!}\mathcal{L}^{2k-3}(A_n) + \tau\gamma_j Q_{n,j}, \quad j = 2, \cdots, k-1, \\ \mathcal{L}^{2k-4}(A_{n+1}) &= \mathcal{L}^{2k-4}(A_n) + \tau\mathcal{L}^{2k-3}(A_n) + \tau\gamma_k \cdot [\![\mathcal{L}^{2k-3}(A_n)]\!], \end{aligned} \tag{3.3.31}$$

where we have

$$\begin{aligned} Q_{n,1} &= V_{n+1} - V_n - \tau A_n - \cdots - \frac{\tau^{2k-2}}{(2k-2)!}\mathcal{L}^{2k-3}(A_n), \\ V_n^{\alpha_1} &= V_n + \tau A_n + \cdots + \frac{\tau^{2k-1}}{(2k-1)!}\mathcal{L}^{2k-3}(A_n) + \alpha_1 Q_{n,1}, \\ Q_{n,j} &= \mathcal{L}^{2j-4}(A_{n+1}) - \mathcal{L}^{2j-4}(A_n) - \cdots - \frac{\tau^{2k-2j+1}}{(2k-2j+1)!}\mathcal{L}^{2k-3}(A_n), \end{aligned} \tag{3.3.32}$$

and for $j = 2, \cdots, k-1$

$$\begin{aligned}
\mathcal{L}^{2j-3}(A_n)^{\alpha_j} &= \mathcal{L}^{2j-3}(A_n) + \cdots + \frac{\tau^{2k-2j+1}}{(2k-2j+1)!}\mathcal{L}^{2k-3}(A_n) + \alpha_j Q_{n,j}, \\
\mathcal{L}^{2k-3}(A_n^{\alpha_k}) &= \mathcal{L}^{2k-3}(A_n) + \alpha_k \cdot [\![\mathcal{L}^{2k-3}(A_n)]\!], \\
\mathcal{L}^{2k-4}(A_n^{\alpha_f}) &= \mathcal{L}^{2k-4}(A_n) + \alpha_f \cdot [\![\mathcal{L}^{2k-4}(A_n)]\!].
\end{aligned} \tag{3.3.33}$$

For $k = 1, 2$, this approach reduces to the second- and third-order generalized-α methods of Section 3.3.3.1, respectively. Next, we define the parameters that control the method's accuracy.

3.3.5 Accuracy analysis

We analyse the accuracy of the method that (3.3.30)-(3.3.33) define by first deriving the amplification matrix. Similarly to the third-order method, we substitute (3.1.20) into (3.3.30) and find a matrix system:

$$L\boldsymbol{U}_{n+1} = R\boldsymbol{U}_n + \mathbf{F}_{\mathbf{n}+\boldsymbol{\alpha}_\mathbf{f}}. \tag{3.3.34}$$

Therefore, similar to (3.1.23), the amplification matrix corresponding to a k-equation system becomes $G = L^{-1}R$ with:

$$G = \begin{bmatrix}
\Lambda_1 & \Xi_{12} & \cdots & \cdots & \Xi_{1k} \\
0 & \Lambda_2 & \Xi_{23} & \cdots & \Xi_{2k} \\
\vdots & & \ddots & & \\
0 & 0 & \cdots & \Lambda_{k-1} & \Xi_{k-1} \\
0 & 0 & \cdots & 0 & \Lambda_k
\end{bmatrix}, \tag{3.3.35}$$

where

$$\Lambda_j = \theta_j \begin{bmatrix} \alpha_j & \alpha_j - \gamma_j \\ -\theta & \alpha_j + (\gamma_j - 1)\theta - 1 \end{bmatrix}, \qquad j = 1, \cdots, k-1, \tag{3.3.36}$$

$$\Lambda_k = \theta_k \begin{bmatrix} \alpha_k + (\alpha_f - 1)\gamma_k\theta & \alpha_k - \gamma_k \\ -\theta & \alpha_k + \alpha_f(\gamma_k - 1)\theta - 1 \end{bmatrix}, \tag{3.3.37}$$

and denoting $\theta_j = (\alpha_j + \gamma_j\theta)^{-1}$, $\theta_k = (\alpha_k + \alpha_f\gamma_k\theta)^{-1}$.

Theorem 3. *The method that equations* (3.3.30)-(3.3.33) *define for semi-discrete system* (3.1.31)*, delivers* $(\frac{3}{2}k)^{th}$ *order of accuracy for even* k *and* $(\frac{3}{2}k + \frac{1}{2})^{th}$ *order for odd* k *in time by setting:*

$$\begin{aligned}
\gamma_j &= \alpha_j - \frac{1}{2}, \qquad \textit{for } j = 1, \cdots, k-1, \\
\gamma_k &= \frac{1}{2} - \alpha_f + \alpha_k.
\end{aligned} \tag{3.3.38}$$

Proof. For a k-system of equations, we expand using Taylor series for k unknowns U^{n+1}, U^{n-1}, $\cdots$, U^{n-k+1} around U^n in time with truncation error of $\mathcal{O}\left(\tau^{\frac{3}{2}k+1}\right)$ for even k, and $\mathcal{O}\left(\tau^{\frac{3}{2}k+\frac{3}{2}}\right)$ for odd k as:

$$\begin{aligned} U^{n+1} &= U^n + \tau V^n + \frac{\tau^2}{2} A^n + \cdots, \\ U^{n-1} &= U^n - \tau V^n + \frac{\tau^2}{2} A^n + \cdots, \\ &\vdots \\ U^{n-k+1} &= U^n - \tau(n-k+1) V^n + \frac{(\tau(n-k+1))^2}{2} A^n + \cdots. \end{aligned} \tag{3.3.39}$$

Then, for second-order accurate schemes, as [56] discusses, the amplification matrix's invariants and the Taylor expansions can be deployed to analyze the accuracy order. The approach applies the Cayley–Hamilton theorem to the resulting 2×2 and 3×3 amplification matrices for parabolic and hyperbolic problems, respectively, (see, e.g., [26, 60]). Herein, for schemes general $k \times k$ amplification matrices, we generalize the analysis to determine the parameters such that the method delivers a desired order of accuracy. Having this in mind, we discuss our general approach in the next section. □

3.3.5.1 Analysing the accuracy of a general system

Herein, we introduce a technique to study the accuracy of a time-marching method with arbitrary order. For this, we first discuss the general form of Cayley–Hamilton theorem. That is, for a general $k \times k$ matrix G, the characteristic polynomial of G, abusing notation, is $p(\lambda) = \det(\lambda I_k - G)$, with I_k denoting the $k \times k$ identity matrix, for which we rewrite the characteristic polynomial $p(\lambda)$ as [53]:

$$p(\lambda) = \lambda^k + c_{k-1}\lambda^{k-1} + \cdots + c_1\lambda^1 + c_0. \tag{3.3.40}$$

Then, instead of the scalar variable λ, one can obtain a similar polynomial to (3.3.40) with the matrix G as:

$$p(G) = G^k + c_{k-1}G^{k-1} + \cdots + c_1 G^1 + c_0 I_k, \tag{3.3.41}$$

for which the Cayley–Hamilton theorem states that the polynomial (3.3.41) equals to the zero matrix, $p(G) = 0$ [55]. We provide further details on the determination of the coefficients c in appendix .1. Next, we multiply (3.3.41) by U^{n-k+1} to have:

$$G^k U^{n-k+1} + c_{k-1}G^{k-1}\, U^{n-k+1} + \cdots + c_1 G^1\, U^{n-k+1} + c_0\, U^{n-k+1} = 0, \tag{3.3.42}$$

given that $U^n = G^n U^0$ [8], we have that:

$$U^{n+1} + c_{k-1}\, U^n + \cdots + c_1\, U^{n-k} + c_0\, U^{n-k+1} = 0. \tag{3.3.43}$$

then, substituting (3.3.39) into (3.3.43) and collecting the terms lead to:

$$(1 + c_{k-1} + \cdots + c0)U^n + (1 - c_{k-2} + \cdots + (-1)^{k+1}c_0)\tau V^n + \cdots = 0. \tag{3.3.44}$$

Using the problem definition, we have $V^n = \lambda_\theta U^n$, $A^n = \lambda_\theta^2 U^n$, which is true for all higher-order terms defined using U^n. Finally, we set the terms γ_j and γ_k to cancel lower-order terms to obtain the optimal accuracy [9].

3.3.5.2 Stability analysis

Follow Sections 3.1.2.2 and 3.3.3.2 closely, we establish the unconditional stability of our method (3.3.30)-(3.3.33); thus, we calculate the eigenvalues of the amplification matrix G in (3.3.35) as:

$$\begin{aligned} 0 = \det\left(G - \tilde{\lambda} I\right) &= \det \begin{bmatrix} \Lambda_1 - \tilde{\lambda} I & \Xi_{12} & \cdots & \cdots & \Xi_{1k} \\ 0 & \Lambda_2 - \tilde{\lambda} I & \Xi_{23} & \cdots & \Xi_{2k} \\ \vdots & & \ddots & & \\ 0 & 0 & \cdots & \Lambda_{k-1} - \tilde{\lambda} I & \Xi_{k-1} \\ 0 & 0 & \cdots & 0 & \Lambda_k - \tilde{\lambda} I \end{bmatrix}, \\ &= \det\left(\Lambda_1 - \tilde{\lambda} I\right) \cdot \det\left(\Lambda_2 - \tilde{\lambda} I\right) \cdots \det\left(\Lambda_k - \tilde{\lambda} I\right). \end{aligned} \tag{3.3.45}$$

Therefore, we bound the spectral radius of each diagonal block to guarantee the overall stability; thus, for $\det\left(\Lambda_j - \tilde{\lambda} I\right) = 0$, $j = 1, \cdots, k$, expression (3.3.22) is valid. Similarly, defining Λ_{lm}^j as the lm component of Λ_j allows us to bound the spectral radius of $\det\left(\Lambda_k^{11} - \tilde{\lambda} I\right) = 0$ by:

$$\gamma_k \theta (1 - 2\alpha_f) \leq 2\alpha_k. \tag{3.3.46}$$

Furthermore, $\det\left(\Lambda_k^{22} - \tilde{\lambda} I - \Lambda_k^{21}\left(\Lambda_k^{11} - \tilde{\lambda} I\right)^{-1} \Lambda_k^{12}\right) = 0$ results in:

$$\left(1 - \gamma_k \theta \theta_k \tilde{\lambda}\right) \cdot \left(1 - (1 + \alpha_f \gamma_k \theta)\theta_k \tilde{\lambda}\right) + (1 - \theta\theta_k) \cdot (1 - (\gamma_k + \alpha_f \gamma_k \theta)\theta_k) = 0. \tag{3.3.47}$$

Therefore, to satisfy (3.3.46) and bound the spectral radius in (3.3.47), we impose the following (for details, see the analysis in the previous section):

$$\alpha_k \geq \alpha_f \geq \frac{1}{2}. \tag{3.3.48}$$

For the other diagonal blocks, $j = 1, \cdots, k-1$, the spectral radius of $\det\left(\Lambda_j^{11} - \tilde{\lambda} I\right) = 0$, is already bounded. Besides, the equations

$$\det\left(\Lambda_j^{22} - \tilde{\lambda} I - \Lambda_j^{21}\left(\Lambda_j^{11} - \tilde{\lambda} I\right)^{-1} \Lambda_j^{12}\right) = 0$$

enforce:

$$\alpha_j \geq 1, \qquad j = 1, \cdots, k-1. \tag{3.3.49}$$

To control the numerical dissipation, we let $\theta \to \infty$ and obtain the eigenvalues of the amplification matrix (3.3.35) as:

$$\begin{aligned} &\lambda_{2j-1} = 0, && \lambda_{2j} = \frac{\gamma_j - 1}{\gamma_j}, \qquad j = 1, \cdots, k-1, \\ &\lambda_{2k-1} = \frac{\alpha_f - 1}{\alpha_f}, && \lambda_{2k} = \frac{\gamma_k - 1}{\gamma_k}. \end{aligned} \tag{3.3.50}$$

Setting $\lambda_{2j} = \rho_j^\infty$ and $\lambda_{2k-1} = \lambda_{2k} = \rho_k^\infty$, the expressions for α_j, α_k, α_f become:

$$\begin{aligned} \alpha_j &= \frac{1}{2}\left(\frac{3 + \rho_1^\infty}{1 + \rho_1^\infty}\right), \qquad j = 1, \cdots, k-1, \\ \alpha_2 &= \frac{1}{2}\left(\frac{3 - \rho_2^\infty}{1 + \rho_2^\infty}\right), \\ \alpha_f &= \frac{1}{1 + \rho_2^\infty}. \end{aligned} \tag{3.3.51}$$

As for the third-order method, choosing $0 \leq \rho_j^\infty, \rho_k^\infty \leq 1$, one controls the dissipation in the high-frequency range while minimizing the dissipation in the low-frequency ones.

Remark 20. *Setting $\rho_1^\infty = \rho_2^\infty = \cdots = \rho_k^\infty = \rho^\infty$, allows us to have a one-parameter family of methods with high accuracy. Additionally, the spectral radius of the system approaches to ρ^∞ in the high-frequency regions.*

Figure 3.8 presents numerical evidence; the method's spectral behaviour for $k \geq 2$ is independent of the accuracy order. Furthermore, compared with the second-order generalized-α method, our generalization improves the spectral behaviour in the mid-frequency regions (e.g., compare the spectral radius for $\rho^\infty = 0.5$). Our method prevents extra damping in these regions with moderate frequency in the second-order generalized-α by approaching the spectral radius to zero for $\rho^\infty = 0.5$.

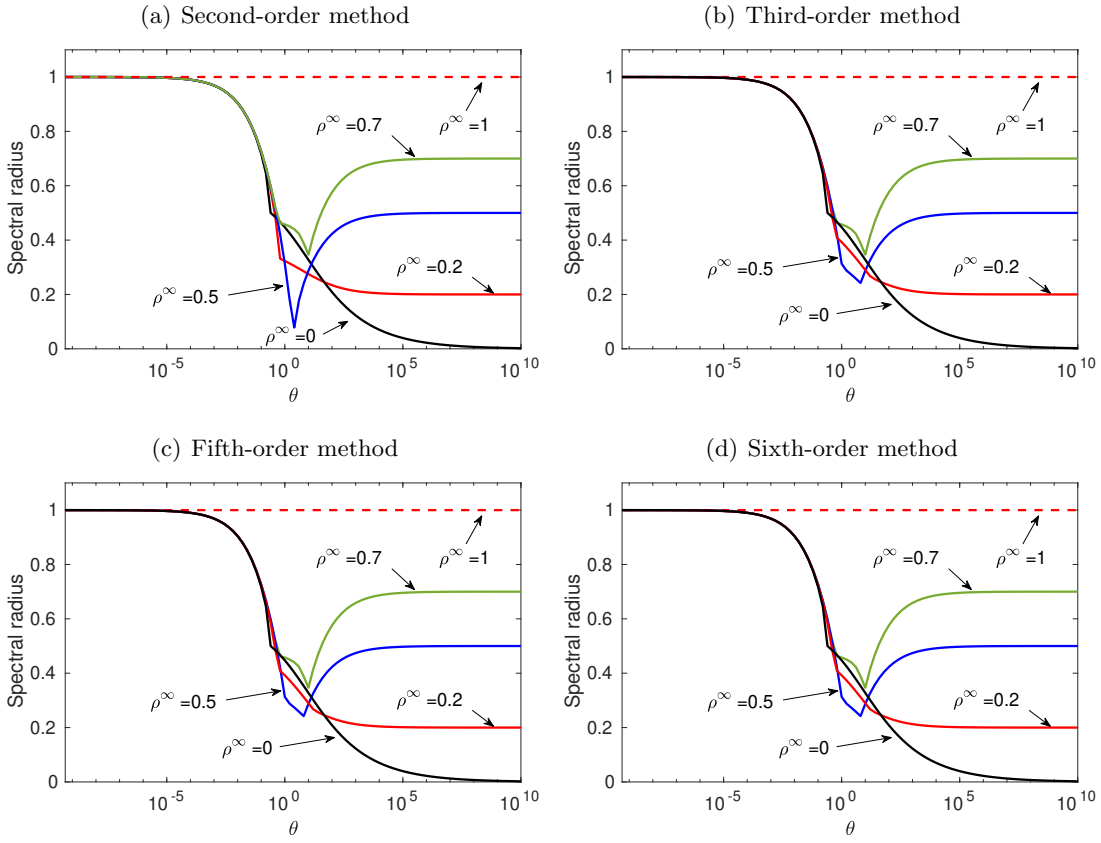


Figure 3.8: Spectral radius of a one-parameter family of methods ($\rho_1^\infty = \rho_2^\infty = \cdots = \rho_k^\infty = \rho^\infty$) [9].

3.3.6 Method's stability for stiff problems

We now investigate the method's ability to solve stiff systems with complex entries (i.e., advection problems). The second Dahlquist barrier states that the stable region of a multistep method for a stiff equation shrinks for accuracy orders higher than two (see, [48]). The method's amplification matrix (3.3.35) requires solving k systems that are form identical to the second-order generalized-α method. Each block decouples from the others; therefore, their eigenvalues are independent as well. Thus, we solve k independent systems that lead to a high-order method with an invariant stability region. While, the analysis in Section 3.3.5.2 supports our claims, herein, we consider a problem with complex eigenvalues, $\lambda_\theta \in \mathbb{C}$. Similarly, we consider the amplification

matrix (3.3.35) to obtain the region of stability as:

$$\left\{\theta \in \mathbb{C} : \left|\frac{\alpha_k + (\alpha_f - 1)\gamma_k\theta}{\alpha_k + \alpha_f\gamma_k\theta}\right| \leq 1,\ \left|\frac{\alpha_j}{\alpha_j + \gamma_j\theta}\right| \leq 1\right\}, \qquad j = 1, \cdots, k-1. \quad (3.3.52)$$

Considering (3.3.48), one can show that imposing $Re(\theta) \geq 0$ is sufficient to satisfy (3.3.52), which proves that our method is A-stable.

Remark 21. *Setting $k = 1$ and $\rho^\infty = 1$, in* (3.3.52) *defines a method with the trapezoidal method's stability region, which has a second-order of accuracy and A-stability.*

Theorem 4. *L-stability: The method introduced in* (3.3.30)-(3.3.32) *shows high-order L-stability for $\rho_\infty = 0$.*

Proof. Given that the method is L-stable (following this section's analysis), recall that $\tilde{\lambda}$ is the eigenvalue of the amplification matrix (3.3.35), thus, we only require to prove [48]:

$$\theta \to \pm\infty \implies \tilde{\lambda} \to 0. \quad (3.3.53)$$

For $\theta \to +\infty$, we already show the spectral behavior in (3.3.50). For $\theta \to -\infty$, we can show that we obtain similar eigenvalues. Therefore, redefining the parameters by setting $\tilde{\lambda} = \rho^\infty$, completes our proof. □

Figure 3.9 shows that the accuracy order and the stability region are independent; the figure shows the boundedness of the system's eigenvalues for a problem with complex eigenvalues. We see that the system's spectral radius behaviour is similar for the second, third, and fifth-order accuracy orders.

3.4 Applications of higher-order generalize-α methods to phase-field models

To complete our discussions on the time-marching methods, we solve with our higher-order methods of Section 3.3 non-linear phase-field models. First, we simulate the folding model of Chapter 2. Thus, considering (2.3.28) and its fully discrete equation (2.5.9), we estimate the non-linear term as:

$$\psi(U_{n+\alpha_f}) \approx \widehat{\psi}(U_{n+\alpha_f}) = \sum_{j=0}^{2k} \mathcal{L}^{(j)}\left(\psi\left(U_{n+1}\right)\right) \frac{(\alpha_f - 1)^j}{j!}, \quad (3.4.1)$$

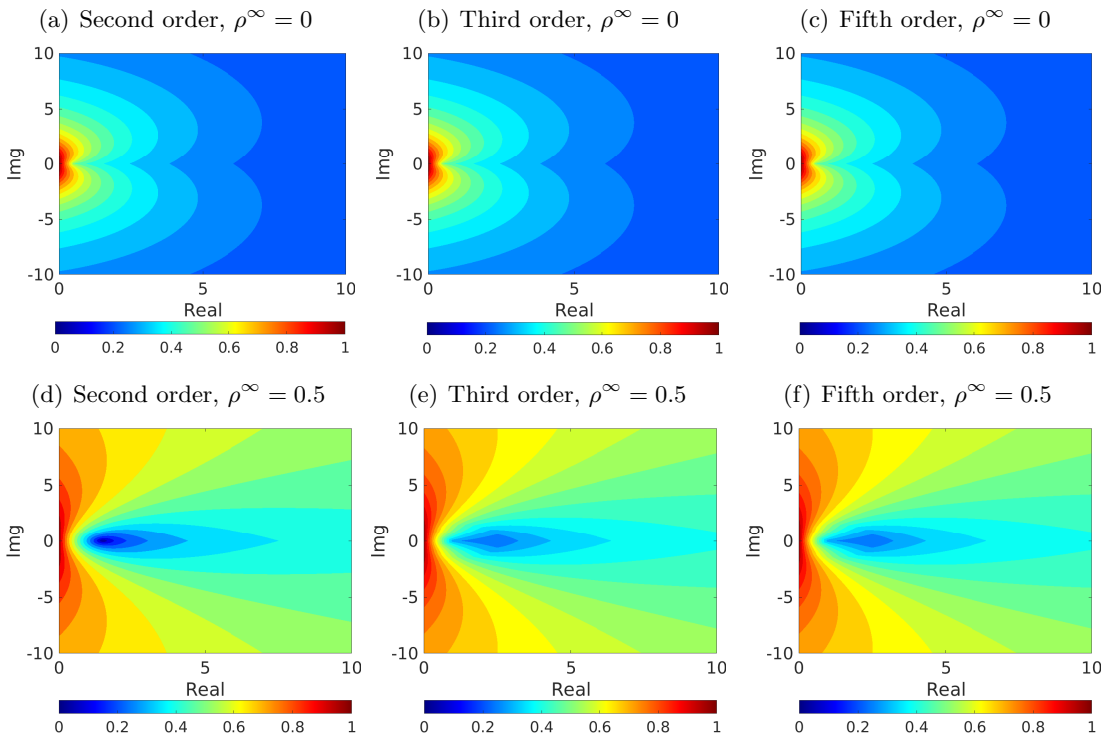


Figure 3.9: Invariant stability behaviour (system's spectral radius unchanged) by increasing order of accuracy [9].

which allows us to introduce a method consisting of solving:

$$
\begin{aligned}
0 &= (\Delta W_{n+1}, \Delta v) - (P\nabla W_{n+1}, \nabla v) + \left(V_n^{\alpha_1} + \widehat{\psi}(W_{n+1}), v\right), \\
0 &= (\Delta \mathcal{L}^{2(j-2)}(A_{n+1}), \Delta v) - (P\nabla \mathcal{L}^{2(j-2)}(A_{n+1}), \nabla v) \\
&\quad + \left(\mathcal{L}^{2j-3}(A_n)^{\alpha_j} + \mathcal{L}^{2(j-2)}(\widehat{\psi}(W_{n+1})), v\right), \\
0 &= (\Delta \mathcal{L}^{2(k-2)}(A_n^{\alpha_f}), \Delta v) - (P\nabla \mathcal{L}^{2(k-2)}(A_n^{\alpha_f}), \nabla v) \\
&\quad + \left(\mathcal{L}^{2k-3}(A_n)^{\alpha_k} + \mathcal{L}^{2(k-2)}(\widehat{\psi}(W_{n+\alpha_k})), v\right),
\end{aligned}
\tag{3.4.2}
$$

with updating terms as in (3.3.31). As Section 3.3 discusses, our method allows the user to control numerical dissipation in the high frequencies, which is an important feature to tackle this problem effectively. Besides, we use the same definitions for the parameters as the ones we describe in Section 3.3 to obtain higher-order accuracy in the temporal domain. For the second-order accuracy in the temporal domain, one can rewrite (3.4.1) as:

$$
\widehat{\psi}(U)_{n+\alpha_f} = \widehat{\psi}(U)_{n+1} + \tau(\alpha_f - 1)\widehat{\dot{\psi}}(U)_{n+1}, \tag{3.4.3}
$$

where

$$\widehat{\dot{\psi}}(U)_{n+1} = V_{n+1} - 3U_{n+1}^2 V_{n+1} + 5\alpha U_{n+1}^4 V_{n+1}.$$

Only $\widehat{\psi}(U)_{n+1}$ is non-linear; thus, the higher-order approximations are linear. For instance, for $j = 2$, we have

$$\begin{aligned}\mathcal{L}^2(\widehat{\psi}(U))_{n+1} = A_{n+1} - 3U_{n+1}^2 A_{n+1} \\ - 6U_{n+1}V_{n+1}^2 + 5\alpha U_{n+1}^4 A_{n+1} + 20\alpha U_{n+1}^3 V_{n+1}^2,\end{aligned} \tag{3.4.4}$$

where the first equation in (3.4.2) determines U_{n+1} and V_{n+1} and the only unknown is A_{n+1} which is linear. Therefore, the solution algorithm, our higher-order time marching method produces, solves a non-linear system and $(k-1)$ linear systems at each time step. Our approach can be easily combined with iterative algorithms.

Remark 22 (Mixed formulation)**.** *In the case of using C^0 finite element methods for the spatial discretization, one can rewrite the* (3.4.2) *using an auxiliary parameter in a mixed form. For this, we need to solve the following matrix problems:*

$$\begin{cases} KW_{n+1} + M\sigma_{n+1} & = 0, \\ MV_n^{\alpha_1} - (M - pK)\sigma_{n+1} + \widehat{\psi}(W_{n+1}) & = 0, \end{cases} \tag{3.4.5}$$

for $j = 2, \cdots, k-1$,

$$\begin{cases} K\mathcal{L}^{2(j-2)}(A_{n+1}) + M\mathcal{L}^{2(j-1)}(\sigma) & = 0, \\ \mathcal{L}^{2j-3}(A_n^{\alpha_j}) - (M - pK)\mathcal{L}^{2(j-1)}(\sigma) + \mathcal{L}^{2(j-1)}(\widehat{\psi}(U))_{n+1} & = 0, \end{cases} \tag{3.4.6}$$

$$\begin{cases} K\mathcal{L}^{2(k-2)}(A_n^{\alpha_f}) + M\mathcal{L}^{2(k-1)}(\sigma) & = 0 \\ \mathcal{L}^{2k-3}(A_n^{\alpha_k}) - (M - pK)\mathcal{L}^{2(k-1)}(\sigma) + \mathcal{L}^{2(k-1)}(\widehat{\psi}(U))_{n+\alpha_k} & = 0 \end{cases} \tag{3.4.7}$$

All parameter definitions and updating expressions are identical.

3.4.1 Numerical results

We approximate the one-dimensional pattern formation case we obtained in (2.3.28). We show our algorithm's energy stability and convergence deploying different spatial and temporal discretizations to solve an example case. The corresponding initial condition is a small perturbation given by $u(x\,,0) = \frac{1}{10^5(1+x^2)}$ on the domain $\Omega = [-40, 40] \in \mathbb{R}$ with given axial force $p = -4$ and $\alpha = 0.3$. To obtain a reference solution to compare the results with that, we perform a simulation with 512 quadratic

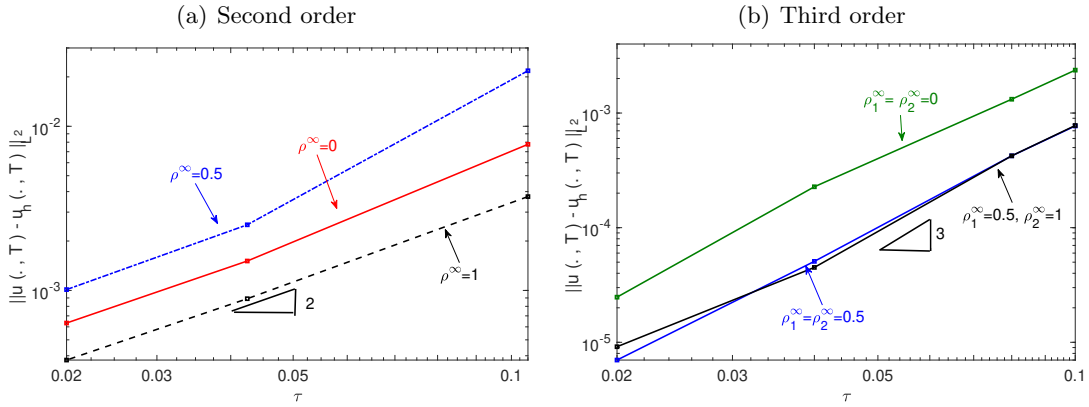


Figure 3.10: Convergence rates for second- and third-order methods

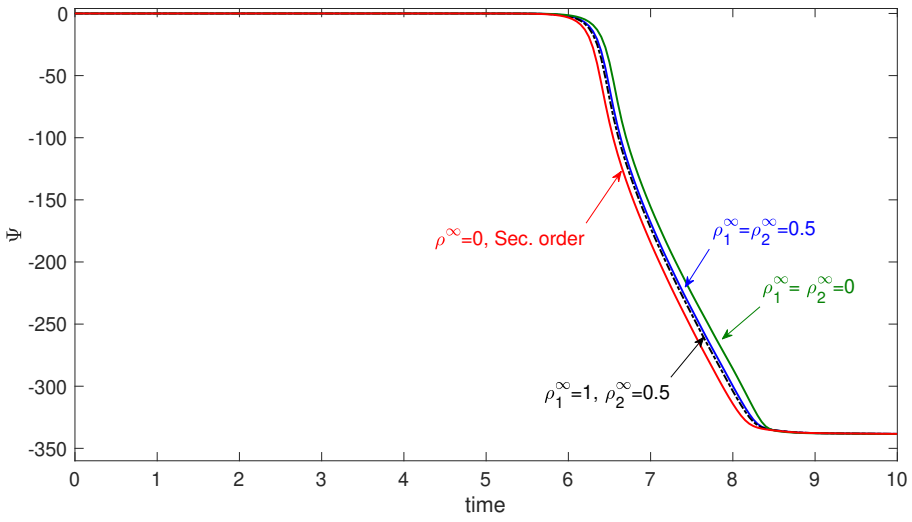


Figure 3.11: Methods' energy behaviour

C^1 elements with time-step of $\tau = 10^{-5}$. In all of the simulations, Newton's method deals with the non-linear term. Figure 3.10 shows the convergence rates for the second- and third-order methods. Also, we show that the convergence rate is independent of ρ^∞'s value. Then, Figure 3.11 illustrates the energy behaviour of our methods, showing that our methods are energy-stable. Finally, figure 3.12 shows the required number of solving a system for different accuracy orders.

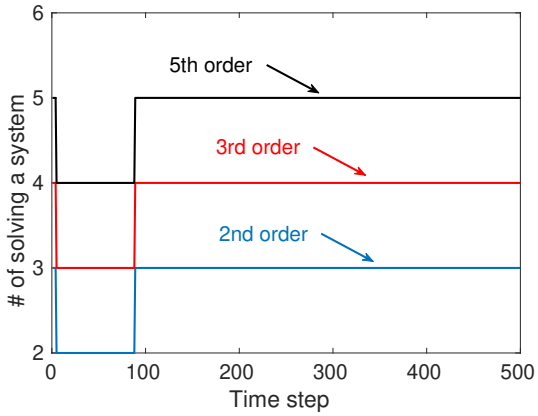


Figure 3.12: Non-linear iterations required by our high-order time-marching methods

3.4.1.1 Phase segregation of block copolymers

In the following example, we simulate a phase-field model describing the phase segregation of block copolymers. The segregation of a solute to surfaces and grain boundaries in solid results in a section of material with a discrete composition and its own set of properties that can affect the overall properties of the material significantly; see, e.g., [41, 68]. Here, we consider diblock copolymers formed by various blocks of two monomers, A and B, connected by covalent junctions. Each monomer tends is repulsed by those of a different kind, and then the copolymer undergoes micro-phase segregation at low temperatures. This process forms A-rich or B-rich regions, which form patterns. In this context, Ohta-Kawasaki theory is advantageous since it does not assume a priori any symmetry of the structure. The Ohta-Kawasaki theory has been widely used to compute the stable states of diblock copolymer melts in both bulk and confinement. The Ohta-Kawasaki model, a modified Cahn-Hilliard equation, reads [75]:

$$\begin{cases} \Delta\phi = \sigma, \\ \dfrac{\partial \phi}{\partial t} = \Delta\eta(\phi) - Gr_2\Delta\sigma - Gr_3(\phi - m), \end{cases} \tag{3.4.8}$$

with the non-linear term

$$\eta(\phi) = \frac{1}{2f}\log\left(\frac{1}{2}(\phi+1)\right) - \frac{1}{2(1-f)}\log\left(\frac{1}{2}(1-\phi)\right) - \frac{1}{2}\chi I\phi + \frac{1}{2f} - \frac{1}{2(1-f)}, \tag{3.4.9}$$

where I is the polymerization's index, χ is the Flory-Huggins interaction parameter, and f is the relative molecular weight. Gr_2 and GR_3 are non-dimensional parameters,

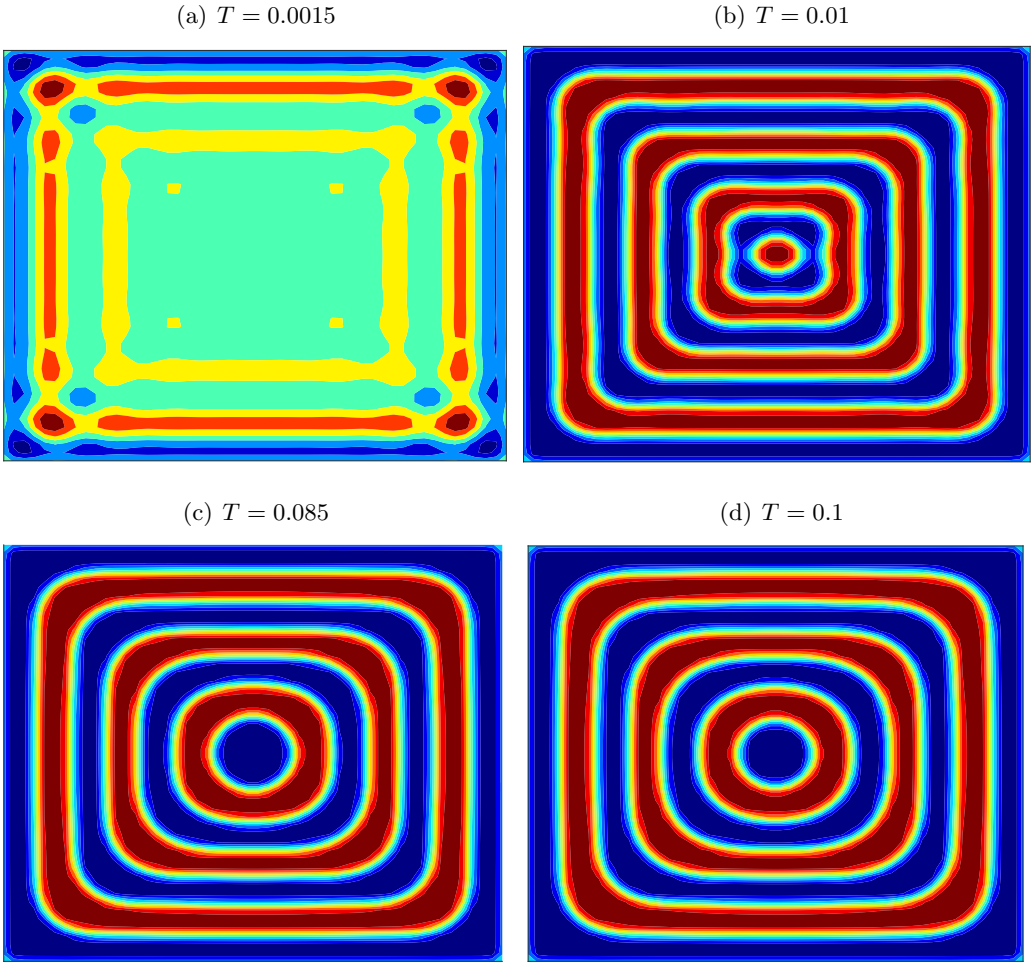


Figure 3.13: Approximate Ohta-Kawasaki solution using the third-order generalized-α

and m is the mass average.

We discretise (3.4.8) using isogeometric analysis in a spatial domain of $\Omega = [-\pi, \pi]$ with number of elements $N = 128 \times 128$. Then, we use our high-order time integrators and set $\tau = 0.0001$. The problem coefficients are also choose as: $GR_2 = 0.024$, $GR_3 = 24$, and $f = 0.5$. Figure 3.13 shows the solution of the system at time 0.0015, 0.01, 0.085, 0.1. Figure 3.14 presents the temporal evolution of the energy of the system as a numerical evident that our approach has energy-stable behaviour.

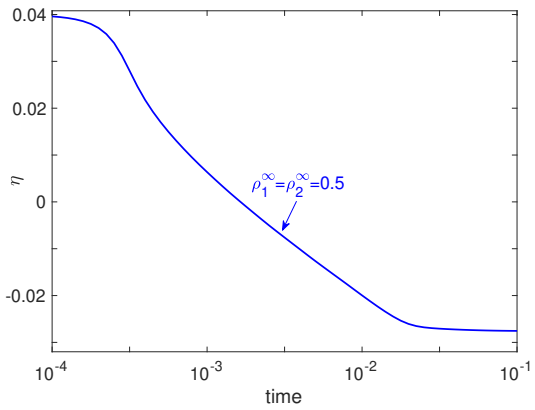


Figure 3.14: Ohta-Kawasaki model's energy evolution approximated using the third-order generalized-α

4

Variational splitting techniques

In this chapter, we introduce splitting techniques that reduce the computational cost of time-dependent problems. We first discretize our problem in space using finite elements on tensor product meshes. In particular, we approximate parabolic or hyperbolic partial differential equations using tensor-product isogeometric analysis (IGA) discretizations in 2D and 3D. Then, we exploit an implicit time marching method such as linear multistep methods or the generalized-α methods to obtain a fully discrete system.

We accelerate the solution of the equation system using a splitting technique that results in a linear cost solver with respect to the total number of degrees of freedom in the system N. Deploying the tensor-product structure of the finite element discretization, we represent the system matrices as Kronecker products of one-dimensional matrices. This reinterpretation of the algebraic system allows us to design simple approximations that deliver linear computational cost $\mathcal{O}(N)$. These approximations build on the ideas first advanced as alternating direction methods [43, 44, 70, 71].

In particular, we develop splitting techniques for the generalized-α method to simulate heat distributions, wave propagation, and linear elasticity problems. Similarly, we extend our approximation framework to two popular families of multistep methods. The s-step Adams-Moulton method [87] that has an $s+1$ order of accuracy in time and the s-step backward differentiation formula (BDF) with accuracy order of s [88]. For example, one may choose higher-order polynomials for spatial discretization. Consequently, to prevent the dominating of error due to temporal discretization, higher-order time integrators (e.g., BDF) should be considered. The increased approximation order raises the computation cost rapidly; thus, splitting techniques are a solution. The main

challenge is to retrieve the optimal convergence in the temporal domain after splitting. Additionally, due to the complex stability behaviour of high-order approaches, it is vital to analyse the effects of splitting on the spectral behaviour of the obtained system. Consequently, we show and present numerical evidence to support our schemes' linear computational cost estimates. Additionally, we prove that the splitting approximations deliver identical spectral properties to the original time integrators.

For explicit methods, the most expensive part of the time marching is the inversion of the mass matrices. Thus, next, we discuss a state-of-the-art approach to precondition the mass matrices resulting from the isogeometric analysis that minimize this cost for complex geometries and single- and multi-patch discretizations.

We detail, analyse, and demonstrate the applicability of the resulting time marching methods in the following three papers:

- Behnoudfar, Pouria, Quanling Deng, and Victor M. Calo. "Split generalized-α method: A linear-cost solver for multi-dimensional second-order hyperbolic systems." Computer Methods in Applied Mechanics and Engineering 376 (2021): 113656 [11].

- Behnoudfar, Pouria, Victor M. Calo, Quanling Deng, and Peter D. Minev. "A variationally separable splitting for the generalized-α method for parabolic equations." International Journal for Numerical Methods in Engineering 121, no. 5 (2020): 828-841 [7].

- Maciej Los, Pouria Behnoudfar, Maciej Paszynski, and Victor M. Calo. "Fast isogeometric solvers for hyperbolic wave propagation problems." Computers & Mathematics with Applications 80, no. 1 (2020): 109-120 [69].

Below, we briefly describe our approaches in the following sections. First, we describe a finite element space built on tensor-product basis functions. Then, building on this simple set of basis functions, we propose our splitting techniques for hyperbolic and parabolic models based on tensor-product basis functions.

4.1 Tensor-product basis functions

Let $\mathcal{T}_h$ be a partition of Ω into non-overlapping tensor-product mesh elements and $K \in \mathcal{T}_h$ be a spatial element with boundary ∂K. Let $(\cdot,\cdot)_S$ be the $L^2(S)$ inner product

with S being a d-dimensional domain (S is typically $\Omega, K, \partial\Omega, \partial K$). Next, to introduce the discrete space associated with the partition $\mathcal{T}_h$, we deploy the Cox-de Boor recursion formula [32, 78] in each direction and then take tensor-product to gain the required basis functions for multiple dimensions.

Nest, assuming $X = \{x_0, x_1, \cdots, x_m\}$ be a knot vector of knots x_j, the j-th B-spline basis function of degree p, denoted as $\theta_j^p(x)$, is determined as [32, 78]

$$\begin{aligned} \theta_j^0(x) &= \begin{cases} 1, & \text{if } x_j \leq x < x_{j+1} \\ 0, & \text{otherwise} \end{cases} \\ \theta_j^p(x) &= \frac{x - x_j}{x_{j+p} - x_j}\theta_j^{p-1}(x) + \frac{x_{j+p+1} - x}{x_{j+p+1} - x_{j+1}}\theta_{j+1}^{p-1}(x). \end{aligned} \tag{4.1.1}$$

The Cox-de Boor recursion formula provides a general knot vector consisting of a set of C^k and p-th order B-spline basis functions, with $p = 1, 2, \cdots$, and $k = 0, 1, \cdots, p-1$. The span of these basis functions form a finite-dimensional subspace of the $H^1(\Omega)$ [21, 39]:

$$V_p^h = \text{span}\{\Theta_j^p\}_{j=1}^{N_h} = \begin{cases} S_k^p = \text{span}\{\theta_j^p(x)\}_{j=1}^{N_x}, & \text{in 1D} \\ S_{k,m}^{p,q} = \text{span}\{\theta_i^p(x)\theta_j^q(y)\}_{i,j=1}^{N_x,N_y}, & \text{in 2D} \\ S_{k,m,n}^{p,q,r} = \text{span}\{\theta_i^p(x)\theta_j^q(y)\theta_l^r(z)\}_{i,j,l=1}^{N_x,N_y,N_z}, & \text{in 3D} \end{cases} \tag{4.1.2}$$

where p, q, r and k, m, n are the approximation and continuity orders. The total number of basis functions in each dimension are denoted by N_x, N_y, N_z.

Next, similar to our previous discussions, the weak formulation of the parabolic and hyperbolic problems straightforwardly read (3.1.29)

$$\begin{aligned} (w_h, \ddot{u}_h) + b(w_h, u_h) &= \ell(w_h), \qquad w_h \in V_h^p(\Omega),\ t > 0, \\ (w_h, \dot{u}_h) + b(w_h, u_h) &= \ell(w_h), \qquad w_h \in V_h^p(\Omega),\ t > 0, \end{aligned} \tag{4.1.3}$$

Constructing the space V_p^h from tensor-product basis functions allows us to derive the bilinear forms $(\cdot,\cdot)$ and $b(\cdot,\cdot)$ that inherit the tensor-product behaviour of the basis. For example, for a 2D case, we have:

$$(\theta_{i_x}^p(x)\theta_{i_y}^q(y), \theta_{j_x}^p(x)\theta_{j_y}^q(y)) = a_x(\theta_{i_x}^p(x), \theta_{j_x}^p(x)) \cdot a_y(\theta_{i_y}^q(y), \theta_{j_y}^q(y)), \tag{4.1.4}$$

where

$$a_x(\theta_{i_x}^p(x), \theta_{j_x}^p(x)) = \int_0^1 \theta_{i_x}^p(x)\theta_{j_x}^p(x)\ \mathrm{d}x, \quad a_y(\theta_{i_y}^q(y), \theta_{j_y}^q(y)) = \int_0^1 \theta_{i_y}^q(y)\theta_{j_y}^q(y)\ \mathrm{d}y \tag{4.1.5}$$

and

$$\begin{aligned} b(\theta^p_{i_x}(x)\theta^q_{i_y}(y), \theta^p_{j_x}(x)\theta^q_{j_y}(y)) = b_x(\theta^p_{i_x}(x), \theta^p_{j_x}(x)) \cdot a_y(\theta^q_{i_y}(y), \theta^q_{j_y}(y)), \\ + a_x(\theta^p_{i_x}(x), \theta^p_{j_x}(x)) \cdot b_y(\theta^q_{i_y}(y), \theta^q_{j_y}(y)), \end{aligned} \tag{4.1.6}$$

where:

$$\begin{aligned} b_x(\theta^p_{i_x}(x), \theta^p_{j_x}(x)) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}x}\theta^p_{i_x}(x)\frac{\mathrm{d}}{\mathrm{d}x}\theta^p_{j_x}(x)\ \mathrm{d}x, \\ b_y(\theta^q_{i_y}(y), \theta^q_{j_y}(y)) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}y}\theta^q_{i_y}(y)\frac{\mathrm{d}}{\mathrm{d}y}\theta^q_{j_y}(y)\ \mathrm{d}y. \end{aligned} \tag{4.1.7}$$

Exploiting the discretization properties in (4.1.4) and (4.1.6), we rewrite mass and stiffens matrices as:

$$\begin{aligned} M &= M^x \otimes M^y, \\ K &= K^x \otimes M^y + M^x \otimes K^y. \end{aligned} \tag{4.1.8}$$

Similarly, we can express the matrices in 3D as:

$$\begin{aligned} M &= M^x \otimes M^y \otimes M^z, \\ K &= K^x \otimes M^y \otimes M^z + M^x \otimes K^y \otimes M^z + M^x \otimes M^y \otimes K^z. \end{aligned} \tag{4.1.9}$$

In this context, we utilize M^ξ and K^ξ with $\xi = x, y, z$ as one-dimensional mass and stiffness matrices, respectively. We call this reformulation as variational separability. These 1D matrices, as well as M and K, are symmetric. The mass matrices are always positive definite, while the stiffness matrices are positive semi-definite; these can be positive definite for some boundary conditions. In the next sections, we use these properties to show the stability of the generalized-α splitting schemes.

Remark 23. *Similarly, we can describe the mass and stiffness matrices using* (4.1.4) *and* (4.1.6) *for any Galerkin discretization based on tensor-product basis functions. For this, it is sufficient to determine the basis functions $\theta^p_j(x)$ using the basis functions associated with the discretization [56].*

4.2 A variational splitting for parabolic problems

We introduce a variational splitting approximation to the fully discrete Galerkin-in-space and generalized-α-in-time discretization of parabolic problems. For this, we use tensor-product grids to generate a splitting method that delivers linear computational cost with respect to the degrees of freedom for multi-dimensional problems. The variationally separable splitting technique builds the variational operators on tensor-product

grids for multi-dimensional problems, which allows us to express the d-dimensional problem as a product of d-systems and error terms.

Let us recall the fully discrete problem obtained using the second-order generalized-α scheme. Substituting (3.3.31) into (3.3.30), we obtain:

$$\alpha_m H[\![V_n]\!] = F_{n+\alpha_f} - KU_n - (M + \tau\alpha_f K)V_n, \tag{4.2.1}$$

where

$$H = M + \eta K \qquad \text{with} \quad \eta = \frac{\tau\gamma\alpha_f}{\alpha_m}. \tag{4.2.2}$$

These splitting methods heavily rely on the following identity:

$$H = M + \eta K = (M_x + \eta K_x) \otimes (M_y + \eta K_y) - \eta^2 K_x \otimes K_y, \tag{4.2.3}$$

where the last term collects terms of order τ^2. Now, let us approximate H as

$$\tilde{H} = (M_x + \eta K_x) \otimes (M_y + \eta K_y) = H - \eta^2 K_x \otimes K_y, \tag{4.2.4}$$

thus,

$$H - \tilde{H} = \eta^2 K_x \otimes K_y. \tag{4.2.5}$$

As a consequence, $\tilde{H}$ is second-order accurate in η approximation to H. Substituting $\tilde{H}$ instead of H in (4.2.1) we propose the generalized-α splitting technique as:

$$\alpha_m (M_x + \eta K_x) \otimes (M_y + \eta K_y)[\![V_n]\!] = F_{n+\alpha_f} - KU_n - (M + \tau\alpha_f K)V_n. \tag{4.2.6}$$

Similarly, in 3D, we approximate H in (4.2.2) by

$$\tilde{H} = (M_x + \eta K_x) \otimes (M_y + \eta K_y) \otimes (M_z + \eta K_z). \tag{4.2.7}$$

Remark 24. *The computational cost of the linear systems with a 1D matrix $M_\xi + \eta K_\xi$, where $\xi = x, y, z$ is linear with respect to the number of degrees of freedom. We detail the procedure in Appendix .2 (see also, [70, 71]).*

Remark 25. *We can introduce a similar splitting to the right-hand side of the matrix multiplying V_n. in* (4.2.6)*. Next, we explain this modification and its impact.*

4.2.1 Other splittings

We split the right-hand-side matrix of (4.2.1). Splitting both sides does not reduce the computational cost further; however, this second splitting improves the approximation's accuracy and its stability.

We denote

$$B = M + \zeta K, \qquad \zeta = \tau\alpha_f. \tag{4.2.8}$$

Now we approximate B using the splitting ideas from the previous section. Thus, in 2D,

$$\tilde{B} = (M_x + \zeta K_x) \otimes (M_y + \zeta K_y). \tag{4.2.9}$$

The 3D splitting follows a similar procedure, see (4.2.7). Thus, we express (4.2.1) as:

$$\alpha_m \tilde{H}[\![V_n]\!] = F_{n+\alpha_f} - KU_n - \tilde{B}V_n. \tag{4.2.10}$$

Alternatively, we can write (4.2.1) as follow

$$\alpha_m H[\![V_n]\!] = F_{n+\alpha_f} - KU_n - \frac{\alpha_m}{\gamma}\big(H + \frac{\gamma - \alpha_m}{\alpha_m}M\big)V_n \tag{4.2.11}$$

Now, we can approximate H using $\tilde{H}$ (4.2.4) in 2D and as (4.2.7) in 3D. This modification results in a second-order accurate scheme in time. We provide numerical evidence in section 4.2.2.3 to show the differences between using the splitting technique (4.2.6) (one side), (4.2.10) (both sides), and (4.2.11) (modified.

4.2.2 Spectral analysis

This section performs the stability analysis and establishes that the splitting schemes are unconditionally stable. We first analyse the standard generalized-α method to sketch how our analysis works. Then, we study the splitting techniques. The generalized-α method is a two-step scheme. The first solves (4.2.1) for $[\![V_n]\!]$. Then, we substitute it in the first equation in (3.3.7) to solve for U_{n+1}, which is the solution at the next time level. Alternatively, supplementing (4.2.1) with the first equation in (3.3.7), we arrive at a matrix formulation of the generalized-α method

$$\begin{bmatrix} U^{n+1} \\ \tau V^{n+1} \end{bmatrix} = \begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m}A^{-1}K & I - \frac{\gamma}{\alpha_m}A^{-1}(M + \tau\alpha_f K) \\ -\frac{\tau}{\alpha_m}A^{-1}K & I - \frac{1}{\alpha_m}A^{-1}(M + \tau\alpha_f K) \end{bmatrix} \begin{bmatrix} U^n \\ \tau V^n \end{bmatrix} + \begin{bmatrix} \frac{\tau\gamma}{\alpha_m}A^{-1}F_{n+\alpha_f} \\ \frac{\tau}{\alpha_m}A^{-1}F_{n+\alpha_f} \end{bmatrix}, \tag{4.2.12}$$

where I is an appropriate identity matrix that matches the dimension; throughout this section, we set $F = 0$ as it does not reduce the generality of the stability analysis.

Remark 26. *In this context, the entries of the matrices obtained from the spatial discretization gain importance as we seek to approximate these matrices using split ones. Therefore, for the analysis, we need to follow a different procedure from the previous sections (e.g., 3.3.3.2). Thus, we compare the spectral behaviour of the systems obtained using the original method and the split one. For this, we consider the matrices H or $\tilde{H}$ for determining the amplification matrix.*

4.2.2.1 The generalized-α method

We analyse the system's stability by spectrally decomposing the matrix K with respect to M (see, e.g., [53]) to have

$$K = MPDP^{-1}, \tag{4.2.13}$$

where D is a diagonal matrix that contains the eigenvalues sorted in ascending order the generalized eigenproblem

$$Kv = \lambda Mv, \tag{4.2.14}$$

and the columns of P are the system's eigenvectors where the j-th column of P corresponds to the eigenvalue $\lambda_j = D_{jj}$. Also, we have $I = PIP^{-1}$. Exploiting (4.2.13) and (4.2.2), we determine

$$\begin{aligned} H^{-1} &= (M + \eta K)^{-1} \\ &= (M + \eta MPDP^{-1})^{-1} \\ &= \left(MP(I + \eta D)P^{-1}\right)^{-1} \\ &= P(I + \eta D)^{-1}P^{-1}M^{-1}. \end{aligned} \tag{4.2.15}$$

Finally, we calculate:

$$\begin{aligned} H^{-1}K &= \left(P(I + \eta D)^{-1}P^{-1}M^{-1}\right)\left(MPDP^{-1}\right) = P(I + \eta D)^{-1}DP^{-1}, \\ H^{-1}M &= \left(P(I + \eta D)^{-1}P^{-1}M^{-1}\right)M = P(I + \eta D)^{-1}P^{-1}. \end{aligned} \tag{4.2.16}$$

Defining $E = (I + \eta D)^{-1}$ allows us to rewrite the amplification matrix in (4.2.12) as:

$$\begin{aligned} \Xi &= \begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m}A^{-1}K & I - \frac{\gamma}{\alpha_m}A^{-1}(M + \tau\alpha_f K) \\ -\frac{\tau}{\alpha_m}A^{-1}K & I - \frac{1}{\alpha_m}A^{-1}(M + \tau\alpha_f K) \end{bmatrix} \\ &= \begin{bmatrix} P & \mathbf{0} \\ \mathbf{0} & P \end{bmatrix} \begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m}ED & I - \frac{\gamma}{\alpha_m}E(I + \tau\alpha_f D) \\ -\frac{\tau}{\alpha_m}ED & I - \frac{1}{\alpha_m}E(I + \tau\alpha_f D) \end{bmatrix} \begin{bmatrix} P^{-1} & \mathbf{0} \\ \mathbf{0} & P^{-1} \end{bmatrix}. \end{aligned} \tag{4.2.17}$$

Thus, we have

$$\begin{bmatrix} U^n \\ \tau V^n \end{bmatrix} = \begin{bmatrix} P & \mathbf{0} \\ \mathbf{0} & P \end{bmatrix} \underbrace{\begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m} ED & I - \frac{\gamma}{\alpha_m} E(I + \tau\alpha_f D) \\ -\frac{\tau}{\alpha_m} ED & I - \frac{1}{\alpha_m} E(I + \tau\alpha_f D) \end{bmatrix}^n}_{\tilde{\Xi}} \begin{bmatrix} P^{-1} & \mathbf{0} \\ \mathbf{0} & P^{-1} \end{bmatrix} \begin{bmatrix} U^0 \\ \tau V^0 \end{bmatrix}. \tag{4.2.18}$$

We establish the method's unconditional stability by bounding the spectral radius of the matrix $\tilde{\Xi}$ by one. For the sake of completeness, we now reproduce the analysis of [9, 60]. We determine the corresponding parameters by presenting the system as (4.4.17). For this, firstly, we the limit $\tau \to 0$. Since D is diagonal, $\tau D \to 0$ and $E \to I$; thus, the matrix $\tilde{\Xi}$ becomes an upper triangular matrix with eigenvalues:

$$\lambda_1 = 1 \qquad \text{and} \qquad \lambda_2 = 1 - \frac{1}{\alpha_m}. \tag{4.2.19}$$

In the other limit of an infinite time step $\tau \to \infty$, one can readily calculate the matrix $\tilde{\Xi}$ as a lower triangular matrix with eigenvalues:

$$\lambda_1 = 1 - \frac{1}{\alpha_f}, \qquad \text{and} \qquad \lambda_2 = 1 - \frac{1}{\gamma}. \tag{4.2.20}$$

For both of these two limiting cases, we arrive at a similar set of conditions as in Section 3.3.5.2. Similarly, in the case of finite time steps, the eigenvalues of the amplification matrix solve:

$$\begin{aligned} 0 &= \det(\tilde{\Xi} - \tilde{\lambda} I) \\ &= \det \begin{bmatrix} \tilde{\Xi}_{11} - \tilde{\lambda} I & \tilde{\Xi}_{12} \\ \tilde{\Xi}_{21} & \tilde{\Xi}_{22} - \tilde{\lambda} I \end{bmatrix} \\ &= \det(\tilde{\Xi}_{11} - \tilde{\lambda} I) \cdot \det\Big(\tilde{\Xi}_{22} - \tilde{\lambda} I - \tilde{\Xi}_{21}(\tilde{\Xi}_{11} - \tilde{\lambda} I)^{-1} \tilde{\Xi}_{12}\Big). \end{aligned} \tag{4.2.21}$$

where

$$\tilde{\Xi} = \begin{bmatrix} \tilde{\Xi}_{11} & \tilde{\Xi}_{12} \\ \tilde{\Xi}_{21} & \tilde{\Xi}_{22} \end{bmatrix} = \begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m} ED & I - \frac{\gamma}{\alpha_m} E(I + \tau\alpha_f D) \\ -\frac{\tau}{\alpha_m} ED & I - \frac{1}{\alpha_m} E(I + \tau\alpha_f D) \end{bmatrix} \tag{4.2.22}$$

is a 2×2 block matrix, with each block being a diagonal matrix. The first part of the eigenvalues of $\tilde{\Xi}$ are defined by the equation $\det(\tilde{\Xi}_{11} - \tilde{\lambda} I) = 0$, where $\tilde{\Xi}_{11} = I - \frac{\tau\gamma}{\alpha_m} ED$ is a diagonal matrix with diagonal entries given by [7]:

$$1 - \frac{\tau \lambda_k \gamma}{\alpha_m + \tau \lambda_k \gamma \alpha_f}, \tag{4.2.23}$$

with λ_k being the $k-$th diagonal entry of D. We guarantee stability by imposing that the absolute value of each of the eigenvalues of $\tilde{\Xi}$ is bounded by one and therefore:

$$-1 \leq 1 - \frac{\tau \lambda_k \gamma}{\alpha_m + \tau \lambda_k \gamma \alpha_f} \leq 1, \tag{4.2.24}$$

or

$$0 \leq \frac{\tau\lambda_k\gamma}{\alpha_m + \tau\lambda_k\gamma\alpha_f} \leq 2. \tag{4.2.25}$$

Since the corresponding parameters are non-negative and D is a positive definite matrix, the left-hand-side inequality is readily satisfied. We reformulate the right-side inequality as:

$$\tau\lambda_k\gamma(1 - 2\alpha_f) \leq 2\alpha_m. \tag{4.2.26}$$

Since $\tau\lambda \geq 0$, the condition $1 - 2\alpha_f \leq 0$ guarantees that the spectral behaviour of $\tilde{\Xi}$ is bounded by one (see, also, Section 3.3.3.2 and [7, 9]. Since each matrix in the second condition in (4.2.21)

$$\det\left(\tilde{\Xi}_{22} - \tilde{\lambda}_k I - \tilde{\Xi}_{21}(\tilde{\Xi}_{11} - \tilde{\lambda}_k I)^{-1}\tilde{\Xi}_{12}\right) = 0, \tag{4.2.27}$$

has a diagonal structure, to calculate the rest of the spectrum of $\tilde{\Xi}$, we solve:

$$\begin{aligned} 0 = &\left(1 - \frac{\tau\lambda_k\gamma}{\alpha_m + \tau\lambda_k\gamma\alpha_f} - \tilde{\lambda}_k\right) \cdot \left(1 - \frac{1 + \tau\lambda_k\alpha_f}{\alpha_m + \tau\lambda_k\gamma\alpha_f} - \tilde{\lambda}_k\right) \\ &+ \left(\frac{\tau\lambda_k}{\alpha_m + \tau\lambda_k\gamma\alpha_f}\right) \cdot \left(1 - \frac{\gamma + \tau\lambda_k\gamma\alpha_f}{\alpha_m + \tau\lambda_k\gamma\alpha_f}\right). \end{aligned} \tag{4.2.28}$$

Thus, we can conclude that a sufficient condition to guarantee the boundedness of the absolute value of the system's eigenvalues by one is:

$$\alpha_m \geq \alpha_f \geq \frac{1}{2}. \tag{4.2.29}$$

4.2.2.2 Stability of the splitting schemes

Herein, we pursuit a spectral study for the splitting technique introduced in section 4.2. Similarly, decomposing the directional matrices K_ξ using M_ξ leads to:

$$K_\xi = M_\xi P_\xi D_\xi P_\xi^{-1}, \tag{4.2.30}$$

where the diagonal D_ξ contains the eigenvalues of the generalized eigenproblem

$$K_\xi v_\xi = \lambda_\xi M_\xi v_\xi \tag{4.2.31}$$

and the columns of P_ξ are the system's eigenvectors. Herein, $\xi = x, y, z$ specify the coordinate directions. We sort the eigenvalues in ascending order and coordinate the

order of D_ξ and the j-th column of P_ξ corresponding to the eigenvalue $\lambda_{\xi,j} = D_{\xi,jj}$. Therefore, for 2D splitting, we use (4.2.30) and (4.2.4) to calculate:

$$\begin{aligned}\tilde{H}^{-1} &= (M_x + \eta K_x)^{-1} \otimes (M_y + \eta K_y)^{-1} \\ &= (M_x + \eta M_x P_x D_x P_x^{-1})^{-1} \otimes (M_y + \eta M_y P_y D_y P_y^{-1})^{-1} \\ &= P_x E_x P_x^{-1} M_x^{-1} \otimes P_y E_y P_y^{-1} M_y^{-1},\end{aligned} \tag{4.2.32}$$

with:

$$E_\xi = (I + \eta D_\xi)^{-1}, \qquad \xi = x, y. \tag{4.2.33}$$

Then, we determine:

$$\begin{aligned}\tilde{H}^{-1}M &= \Big(P_x E_x P_x^{-1} M_x^{-1} \otimes P_y E_y P_y^{-1} M_y^{-1}\Big)(M_x \otimes M_y) \\ &= P_x E_x P_x^{-1} \otimes P_y E_y P_y^{-1} \\ &= \Big(P_x \otimes P_y\Big)\Big(E_x \otimes E_y\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big), \\ \tilde{H}^{-1}K &= \Big(P_x E_x P_x^{-1} M_x^{-1} \otimes P_y E_y P_y^{-1} M_y^{-1}\Big) \\ &\qquad \Big(M_x P_x D_x P_x^{-1} \otimes M_y + M_x \otimes M_y P_y D_y P_y^{-1}\Big) \\ &= P_x E_x D_x P_x^{-1} \otimes P_y E_y P_y^{-1} + P_x E_x P_x^{-1} \otimes P_y E_y D_y P_y^{-1} \\ &= \Big(P_x \otimes P_y\Big)\Big(E_x D_x \otimes E_y + E_x \otimes E_y D_y\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big).\end{aligned} \tag{4.2.34}$$

Defining the following notations

$$\begin{aligned}I &= P_x I_x P_x^{-1} \otimes P_y I_y P_y^{-1} = \Big(P_x \otimes P_y\Big)\Big(I_x \otimes I_y\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big), \\ G &= -\frac{\tau}{\alpha_m}\big(E_x D_x \otimes E_y + E_x \otimes E_y D_y\big),\end{aligned} \tag{4.2.35}$$

Finally, we obtain the blocks of the amplification matrix of the method as:

$$\begin{aligned}\Xi_{11} &= I - \frac{\tau\gamma}{\alpha_m}\tilde{H}^{-1}K = \Big(P_x \otimes P_y\Big)\Big(I_x \otimes I_y + \gamma G\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big), \\ \Xi_{21} &= -\frac{\tau}{\alpha_m}\tilde{H}^{-1}K = \Big(P_x \otimes P_y\Big)G\Big(P_x^{-1} \otimes P_y^{-1}\Big) \\ \Xi_{12} &= I - \frac{\gamma}{\alpha_m}\tilde{H}^{-1}(M + \tau\alpha_f K) \\ &= \Big(P_x \otimes P_y\Big)\Big(I_x \otimes I_y - \frac{\gamma}{\alpha_m}E_x \otimes E_y + \gamma\alpha_f G\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big), \\ \Xi_{22} &= I - \frac{1}{\alpha_m}\tilde{H}^{-1}(M + \tau\alpha_f K) \\ &= \Big(P_x \otimes P_y\Big)\Big(I_x \otimes I_y - \frac{1}{\alpha_m}E_x \otimes E_y + \alpha_f G\Big)\Big(P_x^{-1} \otimes P_y^{-1}\Big).\end{aligned} \tag{4.2.36}$$

and the matrix itself as:

$$\begin{aligned}\Xi &= \begin{bmatrix} I - \frac{\tau\gamma}{\alpha_m}\tilde{H}^{-1}K & I - \frac{\gamma}{\alpha_m}\tilde{H}^{-1}(M + \tau\alpha_f K) \\ -\frac{\tau}{\alpha_m}\tilde{H}^{-1}K & I - \frac{1}{\alpha_m}\tilde{H}^{-1}(M + \tau\alpha_f K) \end{bmatrix} \\ &= \begin{bmatrix} P_x \otimes P_y & \mathbf{0} \\ \mathbf{0} & P_x \otimes P_y \end{bmatrix} \begin{bmatrix} I_x \otimes I_y + \gamma G & I_x \otimes I_y - \frac{\gamma}{\alpha_m}E_x \otimes E_y + \gamma\alpha_f G \\ G & I_x \otimes I_y - \frac{1}{\alpha_m}E_x \otimes E_y + \alpha_f G \end{bmatrix} \\ &\quad \begin{bmatrix} P_x^{-1} \otimes P_y^{-1} & \mathbf{0} \\ \mathbf{0} & P_x^{-1} \otimes P_y^{-1} \end{bmatrix}.\end{aligned} \tag{4.2.37}$$

Accordingly, the spectral radius determines the scheme's stability behaviour:

$$\tilde{\Xi} = \begin{bmatrix} I_x \otimes I_y + \gamma G & I_x \otimes I_y - \frac{\gamma}{\alpha_m}E_x \otimes E_y + \gamma\alpha_f G \\ G & I_x \otimes I_y - \frac{1}{\alpha_m}E_x \otimes E_y + \alpha_f G \end{bmatrix}. \tag{4.2.38}$$

In the limit $\tau \to 0$, we obtain the same sufficient condition as in (4.4.24); similarly, in the limit $\tau \to \infty$, we obtain the eigenvalues $\lambda_1 = 1$, These limits imply that in the limiting case $\tau = \infty$, the method is stable but not A-stable. Following similar arguments as in the unsplit generalized-α method, it is straightforward to prove that the scheme is stable unconditionally (for any finite time step size). The stability behaviour of the splitting on both sides technique follows similar arguments.

Remark 27 (3D splitting)**.** *While the 3D stability analysis is more involved than the 2D analysis, it follows the same logic. Although we omit the derivation, the resulting constraints on α_m, α_f for unconditional stability are unchanged.*

Lastly, we provide error estimations. Our analysis above implies the stability of the splitting schemes. Also, since the splitting error is formally second-order accurate in the temporal domain, providing a sufficiently regular exact solution, one may expect to find the splitting schemes of the second-order accurate in time. Thus, similar to the classical error analysis for parabolic problems, the error estimations read (see, e.g., [96]):

$$\begin{aligned}\|u_h^n - u(t_n)\|_{0,\Omega} &\leq C(u)(h^{p+1} + \tau^2), \\ \|\nabla u(T) - \nabla u_h(T)\|_{0,\Omega} &\leq C(u)(h^p + \tau^2),\end{aligned} \tag{4.2.39}$$

with u_h^n being the estimated solution at time t_n. $C(u)$ denotes a positive constant independent of problem data such as the mesh size h, and time step τ. p is the order of the spatial approximation.

4.2.2.3 Numerical experiments

In this section, numerical evidence demonstrates the performance of the proposed splitting schemes. The simulations verify that the schemes' optimal convergence rates in spatial and temporal domains. Furthermore, the computational cost is linear regarding the total number of degrees of freedom in the multi-dimensional system.
For this, we model (3.1.28) with the exact solution:

$$u = \begin{cases} u(x,y,t) = \sin(\pi x)\sin(\pi y)e^{-2\pi^2 t}, & \text{in} \quad 2D, \\ u(x,y,z,t) = \sin(\pi x)\sin(\pi y)\sin(\pi z)e^{-3\pi^2 t}, & \text{in} \quad 3D, \end{cases} \tag{4.2.40}$$

with corresponding forcing function, boundary and initial data.

Firstly, we show the computational cost of 2D and 3D parabolic problem (3.1.28). For p-th order finite elements or isogeometric elements, the directional mass matrix M_ξ and stiffness matrix $K_\xi, \xi = x, y, z$, have a half-bandwidth p. This is, the matrix $M_\xi + \eta K_\xi$ also has a similar bandwidth. Let K_ξ be m_ξ-dimensional and exploit Gaussian elimination approach. This leads to $\mathcal{O}(p^2 m_\xi)$ operations for solving a linear matrix system $M_\xi + \eta K_\xi$. Therefore, the significant cost of solving (4.2.1) or equivalently (4.2.12) is $\mathcal{O}(p^2 m_x m_y)$ operations for 2D cases and $\mathcal{O}(p^2 m_x m_y m_z)$ operations for 3D problems. Thus, the solution cost of (4.2.1) deploying splitting schemes increases linearly with respect to the degrees of freedom. This feature remains valid for higher-dimensional problems. Furthermore, this allows us for exploiting direct solvers for problems of any dimensions. The reader is referred to [70] for more details on the linear solver.

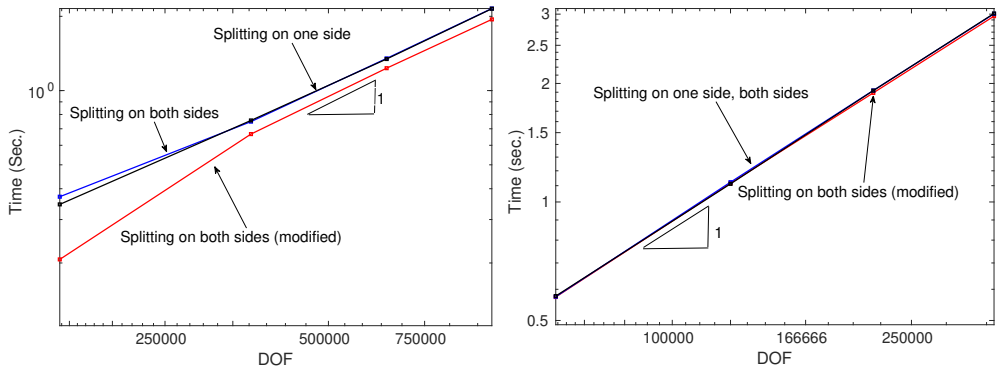


Figure 4.1: Linear computational cost of splitting schemes for C^1 quadratic isogeometric elements with $\tau = 10^{-3}$ and $\rho_\infty = 0$ in 2D (left) and in 3D (right).

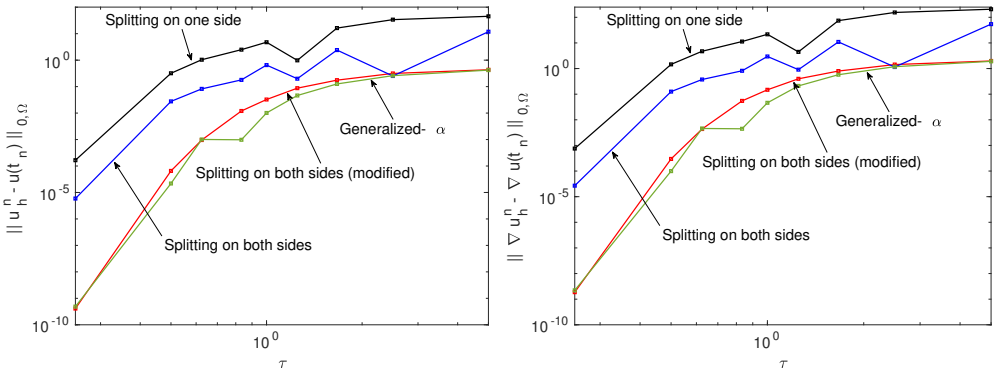


Figure 4.2: Stability validation of splitting schemes when using C^1 quadratic isogeometric elements with final time $T = 5$, $\rho_\infty = 0.5$, and 64×64 uniform elements in 2D.

Figure 4.1 shows that the solution costs for inverting the matrix problems of all splitting schemes are linear with respect to the total number of degrees of freedom in the system for multi-dimensional problems. Herein, we use a direct solver (Gaussian elimination), and as an example, we use C^1 quadratic isogeometric elements for the spatial discretization and a time step size 10^{-3}. However, the cost is also linear when we use either finite elements or isogeometric analysis, which validates the efficiency of the splitting schemes when solving the resulting matrix problems.

Figure 4.2 shows the L^2 norm of u and ∇u errors at the final time $T = 5$ with respect to time step size τ. As τ grows, the errors approach a finite number, which numerically validates the unconditional stability of the generalized-α and splitting schemes. The technique entitled "Splitting on both sides" refers to the scheme given by (4.2.9) while the "Splitting on both sides (modified)" refers to the scheme in equation (4.2.11). Different ρ_∞, mesh configurations, and finite elements with higher-order basis functions produce simulation results with similar behaviours.

Next, we study the spatial convergence rate of the discretization error. The splitting methods deal easily with different spatial discretizations such as finite elements and isogeometric analysis. We consider a 2D test problem and fix the time step size to $\tau = 10^{-4}$. The final time for the simulation is $T = 0.1$. Figure 4.3 shows the errors $\|u(T) - u^h(T)\|_{0,\Omega}$ and $\|\nabla\left(u(T) - u^h(T)\right)\|_{0,\Omega}$ when using C^0 and C^1 quadratic elements. Figure 4.4 shows these errors when using C^2 cubic isogeometric elements. The generalized-α method and all the proposed splitting schemes result in optimal convergence rates for both finite element and isogeometric element discretizations. The error

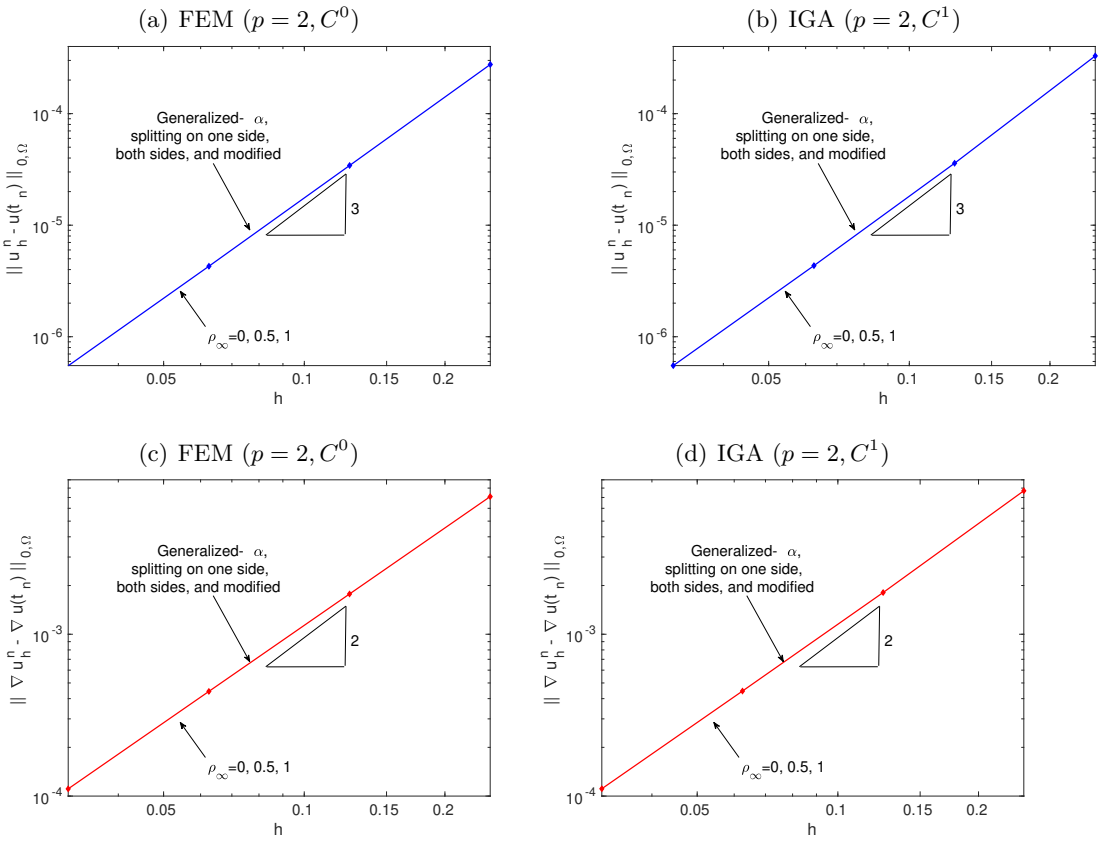


Figure 4.3: L^2 error norm of u & ∇u for C^0 quadratic finite elements and C^1 quadratic isogeometric elements with $\rho_\infty = 0, 0.5, 1$. Final time $t_n = T = 0.1$ & step size $\tau = 10^{-4}$.

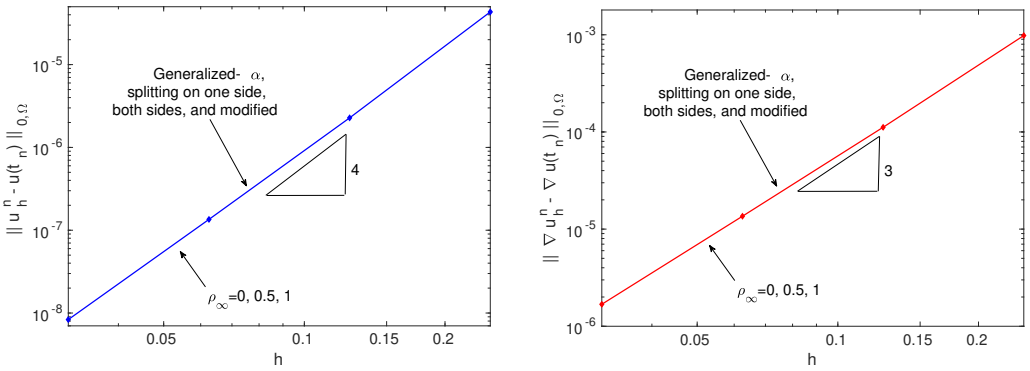


Figure 4.4: L^2 error norm of u and ∇u for C^2 cubic isogeometric elements with $\rho_\infty = 0, 0.5, 1$. Final time $t_n = T = 1$ & step-size $\tau = 10^{-4}$.

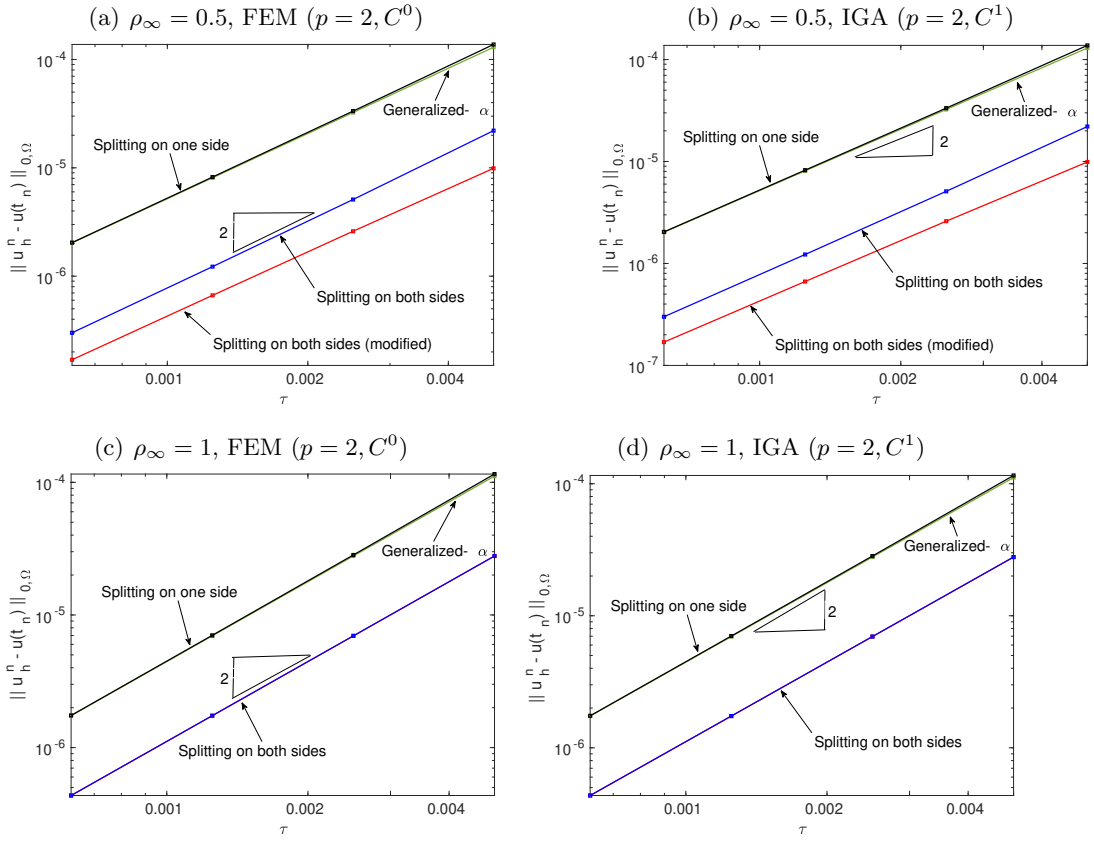


Figure 4.5: L^2 error norm for finite and isogeometric elements on a fixed mesh size $h = 1/64$ and $\rho_\infty = 0.5, 1$. The final time is $t_n = T = 1$.

in all cases converges with the corresponding optimal rates.

We now study the accuracy in time. We consider a 2D test problem on a mesh with 100×100 uniform elements. Figure 4.5 shows the evolution of the solution's L^2 error norm against the time step size τ when using quadratic finite elements and isogeometric analysis. Figure 4.6 shows these errors when using C^2 cubic isogeometric elements. In all cases, the errors converge quadratically, thus verifying the formal estimates of (4.2.39).

The overall solution cost of the matrix system scales linearly with the number of degrees of freedom in the system; thus, in many practical cases, the solution cost is orders of magnitude smaller. We extend these splitting schemes to solve high-order multi-step time integrators in the next section.

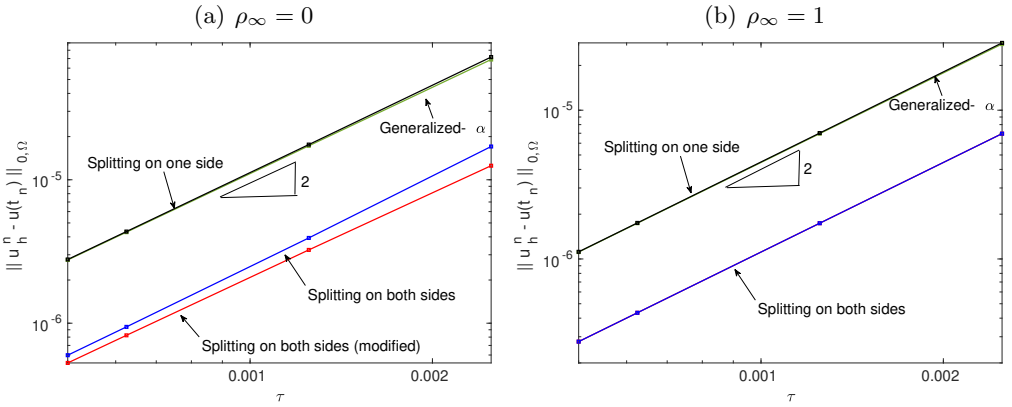


Figure 4.6: L^2 error norm for C^2 cubic isogeometric elements on a fixed mesh size $h = 1/64$ and $\rho_\infty = 0, 1$. The final time is $t_n = T = 1$.

4.3 Variational splitting techniques for high-order linear multistep methods for parabolic equations

Implicit linear multistep methods can approximate first-order stiff differential equations using values from previous time steps. Thereby, the inclusion of information from more points increases the accuracy of the approximation with a limited impact on the computational cost. We use these methods to approximate the semi-discrete matrix problem (3.3.2). We consider a time marching process indexed by n such that $0 = t_0 < t_1 < \cdots < t_N = T$, we approximate U_n, V_n using $U(t_n), \dot{U}(t_n)$, respectively. Similarly, let F_n approximate $F(t_n)$. In general, multistep methods use data from the previous s steps to estimate the solution at the next time step. In particular, a linear multistep method adopts a linear combination of U_n and V_n to obtain a solution value at a specific time. Therefore, a general linear multistep method has the general form [23]

$$U_{n+s} + \alpha_{s-1}\, U_{n+s-1} + \alpha_{s-2}\, U_{n+s-2} + \cdots + \alpha_0\, U_n = \\ = \tau \left(\beta_s\, V_{n+s} + \beta_{s-1}\, V_{n+s-1} + \cdots + \beta_0\, V_n \right),$$

where different coefficient choices for α_i and β_i result in different methods; we apply the splitting techniques to arbitrary combinations of α_i and β_i. We use our variational splitting ideas to the Adams-Moulton and backward differentiation formula (BDF) methods as particular examples. These are widely-used classes of multistep techniques. First, we discuss the *s-step* Adams-Moulton methods and propose our variational splitting

formulations. We then describe the BDF methods and particularize the splitting to this family of methods.

4.3.1 Adams-Moulton methods

Consider a uniform partition of the time interval $[0, T]$ with a time-step size τ such that $t_n = n\tau$. The coefficients are $\alpha_{s-1} = -1$ and $\alpha_{s-2} = ... = \alpha_0 = 0$, while β_i coefficients deliver a method of order $s + 1$. Hence, assuming that the temporal evolution corresponds to an interpolating polynomial of order s, we have [76]

$$\beta_{s-i} = \frac{(-1)^i}{i!(s-i)} \int_0^1 \prod_{j=0, j\neq i}^{s} (u + j - 1)du, \qquad \text{for } i = 0, ..., s. \tag{4.3.1}$$

Let $\mathcal{R}_n = F_n - KU_n$. We adopt the Adams-Moulton methods of order one through five ($s = 0 - 4$). Then, the standard fully discrete formulation becomes

- 2^{nd} order:

$$2M\left(U_{n+1} - U_n\right) = \tau\left(\mathcal{R}_{n+1} + R_n\right) \tag{4.3.2a}$$

- 3^{rd} order:

$$12M\left(U_{n+2} - U_{n+1}\right) = \tau\left(5\mathcal{R}_{n+2} + 8\mathcal{R}_{n+1} - R_n\right) \tag{4.3.2b}$$

- 4^{th} order:

$$24M\left(U_{n+3} - U_{n+2}\right) = \tau\left(9\mathcal{R}_{n+3} + 19\mathcal{R}_{n+2} - 5\mathcal{R}_{n+1} + \mathcal{R}_n\right) \tag{4.3.2c}$$

- 5^{th} order:

$$720M\left(U_{n+4} - U_{n+3}\right) = \tau\left(251\mathcal{R}_{n+4} + 646\mathcal{R}_{n+3} - 264\mathcal{R}_{n+2} + 106\mathcal{R}_{n+1} - 19\mathcal{R}_n\right) \tag{4.3.2d}$$

4.3.1.1 Splitting schemes

Now, we extend ideas from Section 4.2 to propose a splitting for (4.3.2). Thus, we propose the following splitting technique to reduce the solution cost of the linear systems resulting from (4.3.2) for multi-dimension problems:

$$A = M + \eta K = (M_x + \eta K_x) \otimes (M_y + \eta K_y) + \mathcal{O}(\eta^2), \qquad \text{for 2D,} \tag{4.3.3a}$$

$$A = M + \eta K = (M_x + \eta K_x) \otimes (M_y + \eta K_y) \otimes (M_z + \eta K_z) + \mathcal{O}(\eta^2), \qquad \text{for 3D.} \tag{4.3.3b}$$

At each time step, we solve an implicit system and then update vectorial identities explicitly. We factorize the left-hand sides of (4.3.2) using a matrix with the structure $(M + \eta K)$ where η denotes a rational number that depends on the order s. As (4.3.3) shows, the variational splitting introduces an approximation error of order τ^2; thus, to preserve the high approximation order of the multi-step families, we modify the system to solve for an alternative set of variables. To do so, we introduce Taylor series approximations to the right-hand sides of the systems (4.3.2) and define the following approximations:

- 2^{nd} order:

$$U_{n+1} - U_n = \frac{1}{2}\tau\left(V_n + V_{n+1}\right) + \mathcal{O}(\tau^3), \tag{4.3.4a}$$

- 3^{rd} order:

$$U_{n+2} - 2U_{n+1} + U_n = \tau^2\left(V_{n+1}\right)_{,t} + \mathcal{O}(\tau^4), \tag{4.3.4b}$$

- 4^{th} order:

$$U_{n+3} - 3U_{n+2} + 3U_{n+1} - U_n = \frac{1}{2}\tau^3\left(-\left(V_n\right)_{,tt} + 3\left(V_{n+1}\right)_{,tt}\right) + \mathcal{O}(\tau^5), \tag{4.3.4c}$$

- 5^{th} order:

$$U_{n+4} - 4U_{n+3} + 6U_{n+2} - 4U_{n+1} + U_n = \tau^4\left(-\left(V_n\right)_{,ttt} + 2\left(V_{n+1}\right)_{,ttt}\right) + \mathcal{O}(\tau^6), \tag{4.3.4d}$$

where $(\cdot)_{,t}$ represents derivatives with respect to time and V_n approximates the first temporal derivative of $U(t_n)$. Now, we substitute the left-hand terms in (4.3.2) using the linear combinations of known and unknown data derived in (4.3.4). Besides, we use the right-hand side of (4.3.4) to guarantee the accuracy order for our further approximations. Then, substituting these definitions and introducing an auxiliary variable $\mathcal{U}$ results in the following systems that we solve for $\mathcal{U}$:

- 2^{nd} order:

$$\left(2M + \tau K\right)\mathcal{U} = -2\tau K U_n + \tau(F_{n+1} + F_n), \tag{4.3.5a}$$

where $\mathcal{U} = U_{n+1} - U_n$.

- 3^{rd} order:

$$\begin{aligned}(12M + 5\tau K)\,\mathcal{U} = &-12\,(M + \tau K)\,(U_{n+1} - U_n) - 6\tau K\,(U_{n+1} + U_n)\\ &+ \tau\,(5F_{n+2} + 8F_{n+1} - F_n)\,,\end{aligned} \tag{4.3.5b}$$

  where $\mathcal{U} = U_{n+2} - 2U_{n+1} + U_n$.

- 4^{th} order:

$$\begin{aligned}(24M + 9\tau K)\,\mathcal{U} = &-24M\,(2U_{n+2} - 3U_{n+1} + U_n)\\ &- 2\tau K\,(23U_{n+2} - 16U_{n+1} + 5U_n)\\ &+ \tau\,(9F_{n+3} + 19F_{n+2} - 5F_{n+1} + F_n)\,,\end{aligned} \tag{4.3.5c}$$

  where $\mathcal{U} = U_{n+3} - 3U_{n+2} + 3U_{n+1} - U_n$.

- 5^{th} order:

$$\begin{aligned}(720M + 251\tau K)\,\mathcal{U} = &-720M\,(3U_{n+3} - 6U_{n+2} + 4U_{n+1} - U_n)\\ &+ 30\tau K\,(55_{n+3} - 59U_{n+2} + 37U_{n+1} - 9U_n)\\ &+ \tau\,(251F_{n+4} + 646F_{n+3} - 264F_{n+2} + 106F_{n+1} - 19F_n)\end{aligned} \tag{4.3.5d}$$

  where $\mathcal{U} = U_{n+4} - 4U_{n+3} + 6U_{n+2} - 4U_{n+1} + U_n$.

We now apply our variational splittings to these approximations without losing the formal accuracy of the multistep method. That is, the unknown $\mathcal{U}$ on left-hand sides of (4.3.5) are proportional to at least τ^s. Therefore, the resulting approximate expansions in combination with the variational splitting deliver an approximation to $\mathcal{U}$ that is formally of order τ^{s+2} for each multistep method of order $s+1$. As a consequence, the splitting we propose for (4.3.3) can have an arbitrary order of accuracy. Following this logic, we can ignore the higher-order terms in the expansions as they do not modify the formal accuracy of the approximation. In summary, rather than inverting A in (4.3.3), we use the following approximation for 2D

$$\tilde{A} = (M_x + \eta K_x) \otimes (M_y + \eta K_y) \tag{4.3.6}$$

and this one for 3D

$$\tilde{A} = (M_x + \eta K_x) \otimes (M_y + \eta K_y) \otimes (M_z + \eta K_z) \tag{4.3.7}$$

Therefore, substituting $\tilde{A}$ in (4.3.5) leads to a splitting technique with $\frac{\eta}{\tau} = \frac{1}{2}, \frac{5}{12}, \frac{9}{24}, \frac{251}{720}$ that has second to fifth order of accuracy, respectively.

Remark 28. *We solve the equation system* (4.3.5) *with the following strategy. We first solve for* $\mathfrak{U}$ *using the matrix equations of* (4.3.5) *for a given order* s*. With this information, we solve for* U_{n+s-1} *from* $\mathfrak{U}$*'s definition. For example, for the* 3^{rd}*-order method, we solve for* $\mathfrak{U} = U_{n+2} - 2U_{n+1} + U_n$ *using* (4.3.5)*. Then, we update* U_{n+2} *using known information and the following (explicit) identity* $U_{n+2} = \mathfrak{U} + 2U_{n+1} - U_n$*.*

4.3.2 Backward differentiation formulae

Similarly,defining $\mathcal{R}_n = (F_n - KU_n)$, then the general form of a BDF method reads [88]

$$\sum_{j=0}^{s} \alpha_j U_{n+j} = \tau\beta M^{-1}\mathcal{R}_{n+s}, \tag{4.3.8}$$

where α_j and β allow the method to reach the accuracy of order s. For more details on the determination of the corresponding coefficients, see [3, 59]. We adopt standard coefficient definitions for the BDF methods and substitute them in (3.3.2) to obtain:

- 2^{nd} order:

$$M\left(3U_{n+2} - 4U_{n+1} + U_n\right) = 2\tau\mathcal{R}_{n+2}, \tag{4.3.9a}$$

- 3^{rd} order:

$$M\left(11U_{n+3} - 18U_{n+2} + 9U_{n+1} - 2U_n\right) = 6\tau\mathcal{R}_{n+3}, \tag{4.3.9b}$$

- 4^{th} order:

$$M\left(25U_{n+4} - 48U_{n+3} + 36U_{n+2} - 16U_{n+1} + 3U_n\right) = 12\tau\mathcal{R}_{n+4}, \tag{4.3.9c}$$

- 5^{th} order:

$$M\left(137U_{n+5} - 300U_{n+4} + 300U_{n+3} - 200U_{n+2} + 75U_{n+1} - 12U_n\right) = 60\tau\mathcal{R}_{n+5}. \tag{4.3.9d}$$

4.3.2.1 Splitting schemes

As in Section 4.3.1.1, we propose a splitting technique where the solution uses the Taylor expansions of (4.3.4). That is, for (4.3.9), we rewrite the fully discrete equations of (4.3.9). For example, for the BDF methods of orders two to five, we propose

- 2^{nd} order:

$$(3M + 2\tau K)\,\mathcal{U} = (M - 2\tau K)\,U_{n+1} - MU_n + \tau F_{n+2}, \tag{4.3.10a}$$

  where $\mathcal{U} = U_{n+2} - U_{n+1}$.

- 3^{rd} order:

$$(11M + 6\tau K)\,\mathcal{U} = -\,(2M + 6\tau K)\,(2U_{n+2} - U_{n+1}) + 2MU_n + 6\tau F_{n+3}. \tag{4.3.10b}$$

  where $\mathcal{U} = U_{n+3} - 2U_{n+2} + U_{n+1}$.

- 4^{th} order:

$$\begin{aligned}(25M + 12\tau K)\,\mathcal{U} = (9M + 12\tau K)\,(-3U_{n+3} + 3U_{n+2} - U_{n+1})\\ + 3\,(4U_{n+2} - U_n) + 12\tau F_{n+4},\end{aligned} \tag{4.3.10c}$$

  where $\mathcal{U} = U_{n+4} - 3U_{n+3} + 3U_{n+2} - U_{n+1}$.

- 5^{th} order:

$$\begin{aligned}(137M + 60\tau K)\,\mathcal{U} = (62M + 60\tau K)\,(-4U_{n+4} + 6U_{n+3} - 4U_{n+2} + U_{n+1})\\ + 50M\,(3U_{n+3} - 2U_{n+2}) - 12MU_n + 60\tau F_{n+5}\end{aligned}$$

  where $\mathcal{U} = U_{n+5} - 4U_{n+4} + 6U_{n+3} - 4U_{n+2} + U_{n+1}$.

We then use the solution of (4.3.10) to update the value of the unknown. Ad for Adams-Moulton methods, we use the identities in (4.3.4) to introduce a splitting for (4.3.9), where we substitute an approximate expansion that can be inverted in linear cost with formal order of accuracy from second to fifth for $\frac{\eta}{\tau} = \frac{2}{3}, \frac{6}{11}, \frac{12}{25}, \frac{60}{137}$, respectively. We summarize our methods in the algorithm 1.

Algorithm 1 Proposed splitting technique for a 2D problem

Set $T, \tau, M^{\xi}, K^{\xi}, U_0, F_0$, where M^{ξ} and K^{ξ} are the mass an stiffness matrices in ξ direction, where $\xi = x, y$.

for $n = 1$ until $n = \dfrac{T}{\tau}$ **do**,

Set R_n equal to the right-hand side of (4.3.5) for ADM or (4.3.10) for BDF.

Reassemble R_n as a matrix with rows and columns correspond to x and y directions.

Solve $(M_x + \eta K_x)\tilde{\mathcal{U}}_{\backslash} = R_n$ for $\tilde{\mathcal{U}}$

Solve $(M_x + \eta K_x)\mathcal{U}_{\backslash} = \tilde{\mathcal{U}}_{\backslash}$ for $\mathcal{U}_{\backslash}$.

Update the unkown U_{n+s} using $\mathcal{U}_{\backslash}$ and the definitions in (4.3.5) or (4.3.10) for ADM or BDF, respectively.

end for

4.3.3 Spectral analysis

Here, we analyze the splittings' stability and show that the splitting process does not unalter the spectral behaviour of the multistep time marching methods. The conditions that imply the stability of the Adam-Moulton and BDF methods are the same for the original and the split versions. For our investigation, we use the spectral decomposition (4.2.13). Then, A stable method has the modulus of all roots of the characteristic polynomial bounded by one. The boundedness of the roots of the characteristic polynomial for a multistep method with a time-step size τ is [93]:

$$\pi(z; D) = (1 - \tau PDP^{-1}\beta_s)z^s + \sum_{k=0}^{s-1} \left(\alpha_k - \tau PDP^{-1}\beta_k\right) z^k = \rho(z) - \tau PDP^{-1}\sigma(z), \tag{4.3.11}$$

where we define:

$$\rho(z) = z^s + \sum_{k=0}^{s-1} \alpha_k z^k, \qquad \sigma(z) = \sum_{k=0}^{s} \beta_k z^k. \tag{4.3.12}$$

In terms of the polynomials (4.3.11) and (4.3.12), a method delivers order s provided:

$$\rho\left(e^{\tau}\right) - \tau\sigma\left(e^{\tau}\right) = \mathcal{O}\left(\tau^{s+1}\right), \qquad \text{as } \tau \to 0. \tag{4.3.13}$$

Furthermore, appropriate conditions (i.e., CFL conditions [10, 93]) guarantee that all roots of (4.3.11) have their modulus bounded by one. Then, the numerical solution of the multistep method is stable with respect to the round-off error for those values of τ and the particular eigenvalues in D. For brevity, we omit the derivation of the CFL conditions for the BDF and AM methods (for more details, see [3]). We only analyze the impact of the splitting techniques on the stability regions of these well-understood methods; thus, we derive the appropriate characteristic polynomial for the split version.

For the split formulations, we add the following additional terms to (4.3.11):

$$\begin{aligned}
&\text{2D case:} \quad \tau^2(P_x \otimes P_y)(D_x \otimes D_y)(P_x^{-1} \otimes P_y^{-1})\eta(z), \\
&\text{3D case:} \quad \tau^2(I_x \otimes P_y \otimes P_z)(I_x \otimes D_y \otimes D_z + D_x \otimes I_y \otimes D_z + D_x \otimes D_y \otimes I_z) \\
&\qquad\qquad (P_x^{-1} \otimes P_y^{-1} \otimes P_z^{-1})\eta(z),
\end{aligned} \tag{4.3.14}$$

where $\eta(z)$ is defined for order of accuracy two to five as

$$\eta_2(z) = z^2 - z + \mathcal{O}(\tau^3), \tag{4.3.15a}$$

$$\eta_3(z) = z^3 - 2z^2 + z + \mathcal{O}(\tau^4), \tag{4.3.15b}$$

$$\eta_4(z) = z^4 - 3z^3 + 3z^2 - z + \mathcal{O}(\tau^5), \tag{4.3.15c}$$

$$\eta_5(z) = z^5 - 4z^4 + 6z^3 - 4z^2 + z + \mathcal{O}(\tau^6). \tag{4.3.15d}$$

As sections 4.3.1.1 and 4.3.2.1 discuss, the reformulations of the BDF and AM methods with the identities we introduce in (4.3.4) allow us to rewrite (4.3.15) as

$$\begin{aligned}
\eta_2(z) &= \omega_2(z)^2 - \omega_2(z) + \mathcal{O}(\tau^3), \\
\eta_3(z) &= \omega_3(z)^3 - 2\omega_3(z)^2 + \omega_3(z) + \mathcal{O}(\tau^4), \\
\eta_4(z) &= \omega_4(z)^4 - 3\omega_4(z)^3 + 3\omega_4(z)^2 - \omega_4(z) + \mathcal{O}(\tau^5), \\
\eta_5(z) &= \omega_5(z)^5 - 4\omega_5(z)^4 + 6\omega_5(z)^3 - 4\omega_5(z)^2 + \omega_5(z) + \mathcal{O}(\tau^6),
\end{aligned} \tag{4.3.16}$$

where $\omega(z)$'s definition depends on the system's unknown $\mathcal{U}$. Next, using the right-hand sides of the identities (4.3.4), we obtain $\omega(z)$ as

$$\omega_2(z) = \tau z + \frac{1}{2}\tau\,(z)_{,t}, \tag{4.3.17a}$$

$$\omega_3(z) = \tau^2 z, \tag{4.3.17b}$$

$$\omega_4(z) = \tau^3 z - \frac{1}{2}\tau^4\,(z)_{,t}, \tag{4.3.17c}$$

$$\omega_5(z) = \frac{1}{2}\tau^4 z - \tau^5\,(z)_{,t}. \tag{4.3.17d}$$

Given the methods' accuracy, $\tau^2\eta(z)$ has the order of the truncation error and we neglect it when computing the roots of the characteristic polynomial (4.3.11). Therefore, the additional terms in (4.3.14) do not affect the stability regions and, consequently, the CFL conditions for these methods.

Remark 29 (Stability region invariance for split methods)**.** *The split formulation does not change the stability region of the BDF and Adams-Moulton time-marching methods. Nevertheless, the original methods are conditionally stable for any order of greater than two. We believe that the fact that split methods do not modify the stability region of the original methods is an attractive feature of these expansions. Thus, the framework we propose will allow us to split other methods without modifying their stability regions while significantly accelerating the solution of the system.*

4.3.4 Numerical evidence

This section presents the numerical results that validate our proposed splitting methods for different multistep schemes in two and three dimensions. As a test problem, we use a heat transfer problem with zero Dirichlet boundary conditions and an initial condition given by $u(x, y;\, 0) = \sin(\pi x)\sin(\pi y)$. Then, the analytical solution of the problem reads:

$$u(x_1, x_2;\, t) = e^{-2\pi^2 t}\sin(\pi x)\sin(\pi y)\,. \tag{4.3.18}$$

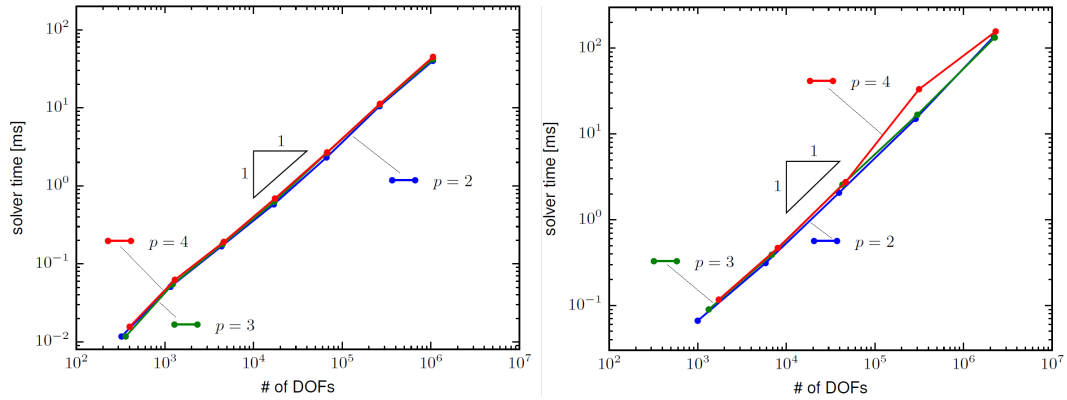


Figure 4.7: Computational cost of split BDF methods for 2D (left) & 3D (right). C^{p-1} isogeometric elements in space.

First, we analyze the computational cost for our splitting schemes for the parabolic problem (3.1.28) in both 2D and 3D. Figures 4.7 and 4.8 show that the costs for solving

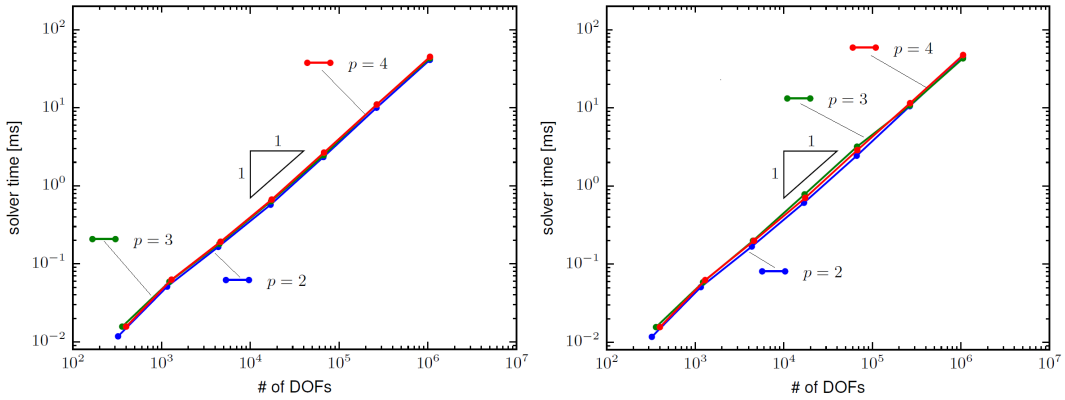


Figure 4.8: Computational cost of split Adam-Moulton methods for 2D (left) & 3D (right). C^{p-1} isogeometric elements in space.

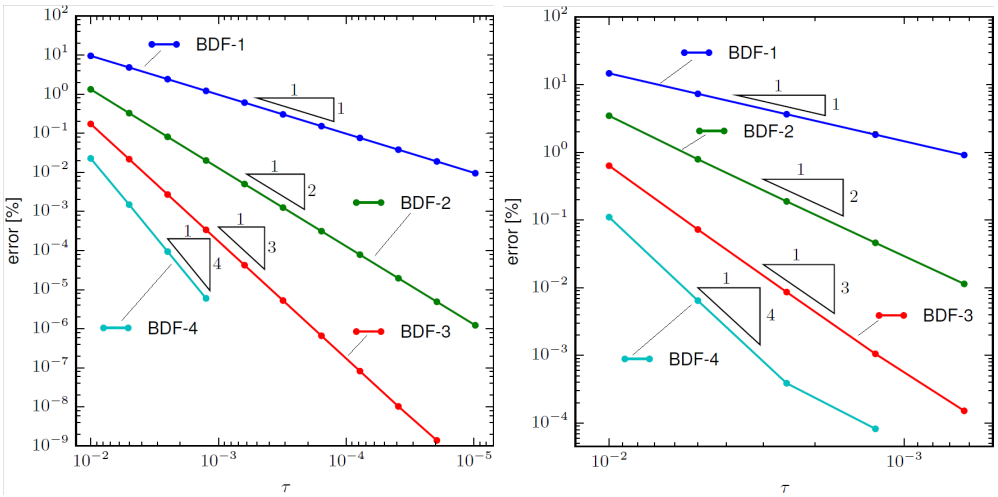


Figure 4.9: L2 error norm for BDF methods with C^4-quintic isogeometric elements 100×100 meshes for 2D (left) & $60 \times 60 \times 60$ meshes for 3D (right).

the resulting algebraic matrix problems for different polynomial approximation orders ($p = 2, 3, 4$) for all BDF and AM methods are linear with respect to the total number of degrees of freedom in the system for multi-dimensional problems.

Next, we present the convergence rates of the different schemes in the time domain. The 2D results are computed over 100×100 mesh. In 3D, we use a $60 \times 60 \times 60$ mesh. The convergence of the BDF-k schemes is of the order $1 : (k)$ the expected one. The convergence of the AM-k schemes is of order $1 : (k + 1)$, as the theory indicates. We obtain that AM-0 scheme converges like 1:1, which is expected. However, the AM-1

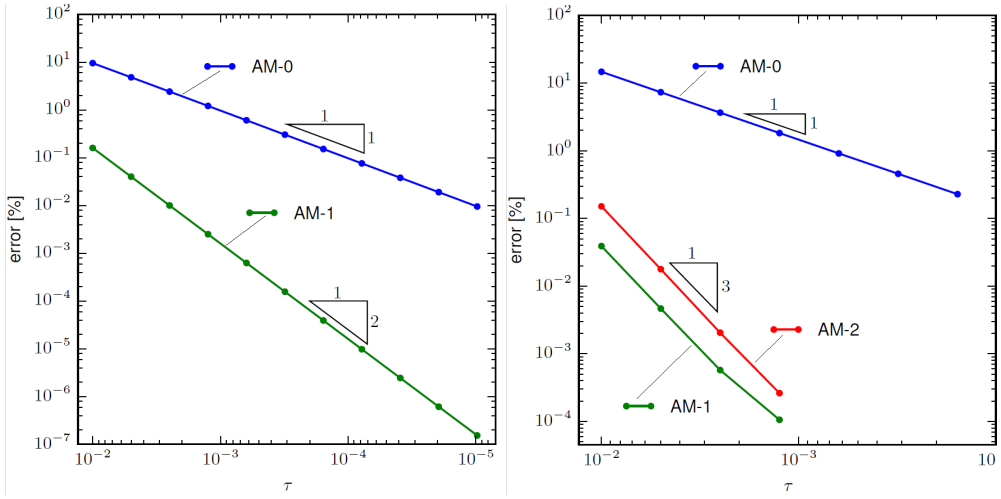


Figure 4.10: L2 error norm for split Adam-Moulton methods with C^4-quintic isogeometric elements 100×100 meshes for 2D (left) & $60 \times 60 \times 60$ meshes for 3D (right).

scheme converges like 1:3, which is faster than expected in this particular case, but this is not an incorrect result.

Remark 30. *Our splitting schemes preserve the stability properties of the original BDF and AM schemes. Therefore, as per the second Dahlquist barrier [31], the linear multistep method can only be A-stable if its order is at most two. Thus, we can only expect unconditional stability for AM-0, AM-1, BDF-1 and BDF-2.*

4.4 A variational splitting technique for hyperbolic problems

In this section, we propose splitting techniques for hyperbolic models. For this, considering the second-order generalized-α method by setting $k = 1$ in (3.1.34), we substitute (3.1.20) into (3.1.34) to have:

$$\begin{aligned}\left(\alpha_m M + \tau^2 \alpha_f \beta K\right) [\![A_n]\!] &= F_{n+\alpha_f} - \left[MA_n + K\left(U_n + \tau\alpha_f V_n + \frac{\tau^2}{2}\alpha_f A_n\right)\right] \\ &= F_{n+\alpha_f} - \left(M + \frac{\tau^2}{2}\alpha_f K\right) A_n - \tau\alpha_f K V_n - K U_n. \quad (4.4.1)\end{aligned}$$

Then, we rewrite (4.4.1) as

$$\alpha_m G[\![A_n]\!] = F_{n+\alpha_f} - \left(M + \frac{\tau^2}{2}\alpha_f K\right) A_n - \tau\alpha_f K V_n - K U_n, \quad (4.4.2)$$

where

$$G = M + \eta K, \qquad \eta = \frac{\tau^2 \alpha_f \beta}{\alpha_m}. \tag{4.4.3}$$

Thus, using tensor-product meshes in 2D and (4.1.4) and (4.1.6), we have (for more details, see, section 4.2 and [11]):

$$G = (M^x + \eta K^x) \otimes (M^y + \eta K^y) - T_\eta, \tag{4.4.4}$$

where

$$T_\eta = \eta^2 K^x \otimes K^y = \mathcal{O}(\eta^2). \tag{4.4.5}$$

Therefore, T_η is of order $\mathcal{O}(\tau^4)$ and allows us to split the operator G as:

$$\tilde{G} = (M^x + \eta K^x) \otimes (M^y + \eta K^y). \tag{4.4.6}$$

Besides, we substitute K with an approximation derived using similar logic. We derive this equivalent expression using (4.4.3) as (see, our detailed discussions in 4.2.1):

$$K = \frac{1}{\eta} G - \frac{1}{\eta} M. \tag{4.4.7}$$

Thus, our approximation reads:

$$K \approx \frac{1}{\eta} \left(\tilde{G} - M \right). \tag{4.4.8}$$

Finally, we rewrite (4.4.2) as

$$\alpha_m \tilde{G} [\![A_n]\!] = F_{n+\alpha_f} - M A_n - \frac{1}{\eta} \left(\tilde{G} - M \right) \left[\frac{\tau^2 \alpha_f}{2} A_n + \tau \alpha_f V_n + U_n \right]. \tag{4.4.9}$$

We split the 3D system using a similar argument. This modification is second-order accurate in the temporal domain and reduces the matrix-assembly cost of the system as we only need $\tilde{G}$. Besides, this modification provides more accurate estimations (see, numerical examples). Additionally, the system in (4.4.9) delivers the user-control on the numerical dissipation. Algorithm 2 summarizes the method for 2D problems. In the next section, we analyze the stability of the method and present its advantages.

Remark 31. *Similar to our techniques for parabolic problems, the computational cost of the 1D matrix system* $(M_\xi + \eta K_\xi)^{-1}$*, is linear with respect to the number of degrees of freedom [70, 71]. Therefore, the total cost for the system system* (4.4.9) *for a multi-dimensional problem is linear.*

Algorithm 2 A linear-cost split generalized-alpha method

Set $T, \tau, \rho_\infty, M^\xi, K^\xi, U_0, V_0, F_0$, where M^ξ and K^ξ are the mass an stiffness matrices in ξ direction with $\xi = x, y$.

for $n = 1$ until $n = \dfrac{T}{\tau}$ **do**,

Set $R_n = F_{n+\alpha_f} - MA_n - \dfrac{1}{\eta}\left(\tilde{G} - M\right)\left[\dfrac{\tau^2\alpha_f}{2}A_n + \tau\alpha_f V_n + U_n\right]$.

Reassemble R_n as a matrix representation, which rows and columns belong to values corresponding to x and y directions.

Solve $\alpha_m(M^x + \eta K^x)[\![\tilde{A}_x]\!] = R_n$ for $[\![\tilde{A}_x]\!]$,

Solve $(M^y + \eta K^y)[\![A_n]\!] = [\![\tilde{A}_x]\!]$ for $[\![A_n]\!]$.

Update V_{n+1} from A_{n+1} using (3.2.1b).

Update U_{n+1} form V_{n+1} and A_{n+1} using (3.2.1c).

end for

4.4.1 Spectral analysis

We closely follow the approach in [11, 26]; we first study the standard generalized-α method and then analyze our splitting method.

4.4.1.1 The generalized-α method

The classical generalized-α method (4.4.2) results in the following:

$$\begin{bmatrix} U^{n+1} \\ \tau V^{n+1} \\ \tau^2 A^{n+1} \end{bmatrix} = \Xi \begin{bmatrix} U^{n} \\ \tau V^{n} \\ \tau^2 A^{n} \end{bmatrix}, \tag{4.4.10}$$

where

$$\Xi = \begin{bmatrix} I - \tau^2{\alpha_f}^2 J & I - \tau^2\alpha_f^3 J & \dfrac{1}{2}I - \alpha_f^2 R \\ -\tau^2\gamma J & I - \tau^2\gamma\alpha_f J & I - \gamma R \\ -\tau^2 J & -\tau^2\alpha_f J & I - R \end{bmatrix}, \tag{4.4.11}$$

and I is the identity matrix. We also denote

$$J = (\alpha_m G)^{-1} K, \qquad R = (\alpha_m G)^{-1}\left(M + \frac{\tau^2}{2}\alpha_f K\right), \tag{4.4.12}$$

Then, using the spectral decomposition eq:sd and (4.4.3), we determine:

$$G^{-1} = (M + \eta K)^{-1} = \left(M + \eta MPDP^{-1}\right)^{-1} = \left(MP(I + \eta D)P^{-1}\right)^{-1}$$
$$\therefore \qquad G^{-1} = P\,(I + \eta D)^{-1}\,P^{-1}M^{-1}, \tag{4.4.13}$$

which allows us to obtain

$$G^{-1}K = \left(P(I + \eta D)^{-1}P^{-1}M^{-1}\right)\left(MPDP^{-1}\right)$$
$$\therefore \qquad G^{-1}K = P\,(I + \eta D)^{-1}\,DP^{-1}, \tag{4.4.14}$$
$$G^{-1}M = \left(P(I + \eta D)^{-1}P^{-1}M^{-1}\right)M$$
$$\therefore \qquad G^{-1}M = P(I + \eta D)^{-1}P^{-1}. \tag{4.4.15}$$

We also define $E = (I + \eta D)^{-1}$ and

$$\Pi = \begin{bmatrix} P & 0 & 0 \\ 0 & P & 0 \\ 0 & 0 & P \end{bmatrix}.$$

With this in mind, using (4.4.11), we rewrite the amplification matrix Ξ in (4.4.10) as:

$$\Xi = \Pi \begin{bmatrix} I - \tau^2 \dfrac{{\alpha_f}^2}{\alpha_m} ED & I - \tau^2 \dfrac{{\alpha_f}^3}{\alpha_m} ED & \dfrac{1}{2} I - \dfrac{{\alpha_f}^2 E}{\alpha_m}\left(I + \dfrac{\tau^2}{2}\alpha_f D\right) \\ -\tau^2 \dfrac{\gamma}{\alpha_m} ED & I - \tau^2 \gamma \dfrac{\alpha_f}{\alpha_m} ED & I - \dfrac{\gamma E}{\alpha_m}\left(I + \dfrac{\tau^2}{2}\alpha_f D\right) \\ -\dfrac{\tau^2}{\alpha_m} ED & -\tau^2 \dfrac{\alpha_f}{\alpha_m} ED & I - \dfrac{E}{\alpha_m}\left(I + \dfrac{\tau^2}{2}\alpha_f D\right) \end{bmatrix} \Pi^{-1}, \tag{4.4.16}$$

which introducing $\Xi = \Pi^{-1}\,\tilde{\Xi}\,\Pi$ leads to

$$\begin{bmatrix} U^n \\ \tau V^n \\ \tau^2 A^n \end{bmatrix} = \Pi\,\tilde{\Xi}^n \Pi^{-1} \begin{bmatrix} U^0 \\ \tau V^0 \\ \tau^2 A^0 \end{bmatrix}. \tag{4.4.17}$$

Thus, the method behaves unconditionally stable if the spectral radius of either of the similar matrices Ξ or $\tilde{\Xi}$ is bounded by one. Herein, we omit the analysis for brevity and state that the method is unconditionally stable for specific values of α_m and α_f given in (3.1.18) (see our detailed discussions in 4.2 and [7]).

4.4.1.2 Stability of the splitting scheme

Here, we analyze the proposed splitting approach. Similarly, we spectrally decompose the stiffness matrix K_ξ with respect to the mass matrix M_ξ in each direction to obtain (4.2.30). Then, for the 2D case, we have

$$\begin{aligned} \tilde{G}^{-1} &= P_x E_x P_x^{-1} M_x^{-1} \otimes P_y E_y P_y^{-1} M_y^{-1}, \\ \tilde{G}^{-1} M &= (P_x \otimes P_y)(E_x \otimes E_y)\left(P_x^{-1} \otimes P_y^{-1}\right), \end{aligned} \tag{4.4.18}$$

where:

$$E_\xi = (I + \eta D_\xi)^{-1}, \qquad \xi = x, y. \tag{4.4.19}$$

Now, using the following identity:

$$I = P_x I_x P_x^{-1} \otimes P_y I_y P_y^{-1} = (P_x \otimes P_y)(I_x \otimes I_y)\left(P_x^{-1} \otimes P_y^{-1}\right), \tag{4.4.20}$$

we calculate the blocks of the amplification matrix $\tilde{\Xi}$ as:

$$\tilde{\Xi}_{11} = I_x \otimes I_y - \frac{1}{\alpha_f}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21a}$$

$$\tilde{\Xi}_{12} = E_x \otimes E_y, \tag{4.4.21b}$$

$$\tilde{\Xi}_{13} = \left(\frac{1}{2} - \frac{\beta}{\alpha_m}\right) E_x \otimes E_y, \tag{4.4.21c}$$

$$\tilde{\Xi}_{21} = -\frac{\gamma}{\beta\alpha_f}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21d}$$

$$\tilde{\Xi}_{22} = I_x \otimes I_y - \frac{\gamma}{\beta}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21e}$$

$$\tilde{\Xi}_{23} = I_x \otimes I_y - \frac{\gamma}{\alpha_m}(E_x \otimes E_y) - \frac{\gamma}{2\beta}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21f}$$

$$\tilde{\Xi}_{31} = -\frac{1}{\beta\alpha_f}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21g}$$

$$\tilde{\Xi}_{32} = -\frac{1}{\beta}(I_x \otimes I_y - E_x \otimes E_y), \tag{4.4.21h}$$

$$\tilde{\Xi}_{33} = I_x \otimes I_y - \frac{1}{\alpha_m}(E_x \otimes E_y) - \frac{1}{2\beta}(I_x \otimes I_y - E_x \otimes E_y). \tag{4.4.21i}$$

By denoting $\tilde{E} = E_x \otimes E_y$, we obtain:

$$\tilde{\Xi} = \begin{bmatrix} I - \frac{1}{\alpha_f}\left(I - \tilde{E}\right) & \tilde{E} & \left(\frac{1}{2} - \frac{\beta}{\alpha_m}\right)\tilde{E} \\ -\frac{\gamma}{\beta\alpha_f}\left(I - \tilde{E}\right) & I - \frac{\gamma}{\beta}\left(I - \tilde{E}\right) & I - \frac{\gamma\tilde{E}}{\alpha_m} - \frac{\gamma}{2\beta}(I - \tilde{E}) \\ -\frac{1}{\beta\alpha_f}\left(I - \tilde{E}\right) & -\frac{1}{\beta}\left(I - \tilde{E}\right) & I - \frac{\tilde{E}}{\alpha_m} - \frac{1}{2\beta}(I - \tilde{E}) \end{bmatrix}. \tag{4.4.22}$$

The spectral behaviour of the scheme follows the same logic as analysis of the generalized-α method, where we compute the spectral radius of (4.4.22). First, in the limit $\sigma := \tau^2 D_\xi \to 0$, since D_ξ is diagonal, $E_\xi \to I$ and consequently, we have $\tilde{E} \to I$. Hence, $\tilde{\Xi}$ becomes upper triangular, and its eigenvalues are:

$$\lambda_1 = \lambda_2 = 1, \qquad \lambda_3 = 1 - \frac{1}{\alpha_m}. \tag{4.4.23}$$

To bound $|\lambda_3|$ by one, we obtain the condition:

$$\alpha_m \geq \frac{1}{2}. \tag{4.4.24}$$

Likewise, for $\sigma \to \infty$, we have $\tilde{E} \to \mathbf{0}$, and therefore:

$$\tilde{\Xi} = \begin{bmatrix} I - \dfrac{I}{\alpha_f} & \mathbf{0} & \mathbf{0} \\ -\dfrac{\gamma I}{\beta \alpha_f} & I - \dfrac{\gamma I}{\beta} & I - \dfrac{\gamma I}{2\beta} \\ -\dfrac{I}{\beta \alpha_f} & -\dfrac{I}{\beta} & I - \dfrac{I}{2\beta} \end{bmatrix}. \tag{4.4.25}$$

Setting γ as in (3.1.25), the system's eigenvalues remain bounded by one unconditionally. Furthermore, following [26], we introduce β such that the complex part of the eigenvalues of (4.4.25) vanishes and express the parameters α_m and α_f in terms of the spectral radius ρ_∞ similar to (3.1.27). This allows us to control the high-frequency dissipation.

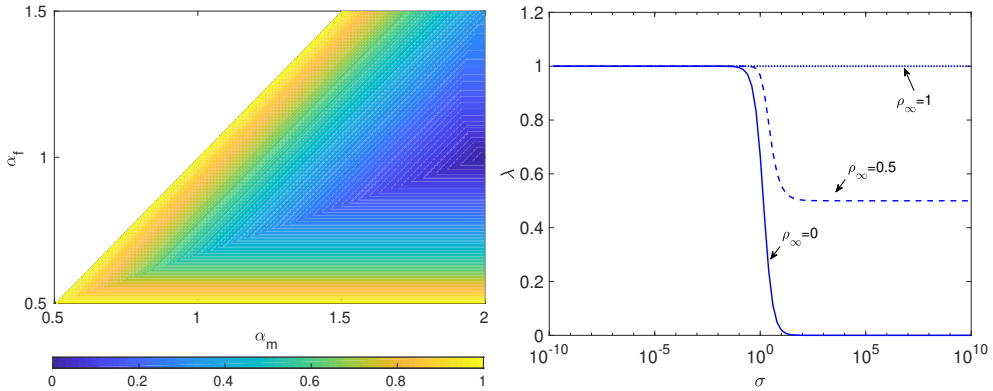


Figure 4.11: The stability region of the split system for different combinations of α_f and α_m at $\sigma \to \infty$ (left) and dissipation control in high-frequency regions (right).

Figure 4.11 presents the region where the split system (4.4.25) has bounded eigenvalues for various α_m and α_f. This figure also shows the system's spectral behaviour in the medium frequency region (i.e., finite time-steps).

Remark 32. *The derivation of the conditions on α_m, α_f for unconditional stability for the 3D splitting is more involved but follows the same logic; thus, we omit those details.*

4.4.2 Numerical examples

We provide numerical experiments to evidence our analysis and demonstrate the splitting methods' efficiency. In all simulations, our solutions converge with optimal rates in the spatial and temporal domains. Furthermore, we present a linear computational cost of the splitting schemes with respect to the total number of degrees of freedom in the system and the dimensionality of the physical problem.

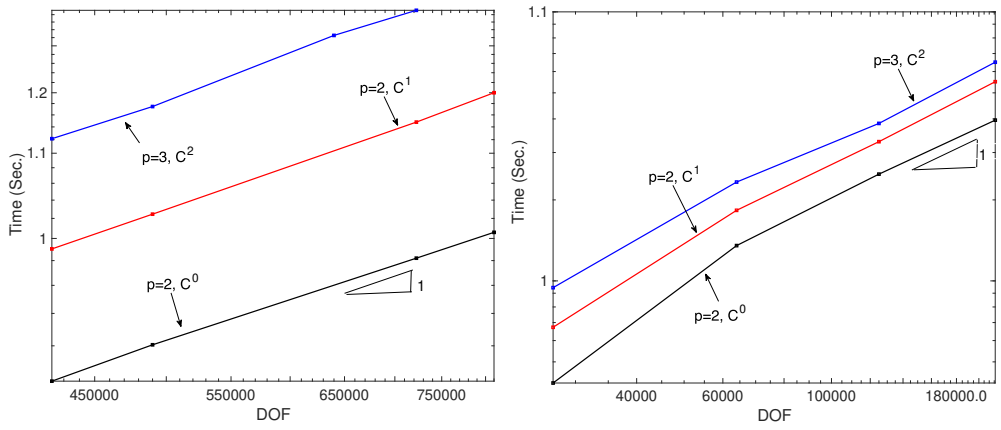


Figure 4.12: Linear computational cost of splitting scheme for C^0 & C^1 quadratics and C^2 cubics with $\rho_\infty = 0.5$ for 2D (left) & 3D (right).

4.4.2.1 Test with exact solution

Herein, we model the problem (3.1.28) with the given forcing term, boundary, and initial data derived from the exact solution:

$$\begin{cases} u(x,y,t) = \sin(\pi x)\sin(\pi y)\left(\sin(\sqrt{2}\pi t) + \cos(\sqrt{2}\pi t)\right), & \text{in} \quad 2D, \\ u(x,y,z,t) = \sin(\pi x)\sin(\pi y)\sin(\pi z)\left(\sin(\sqrt{3}\pi t) + \cos(\sqrt{3}\pi t)\right), & \text{in} \quad 3D. \end{cases} \tag{4.4.26}$$

First, we show that our method has a linear computation cost. The inversion of the matrix is the major cost of simulating (4.4.9); the number of operations for a 1D problem reads $\mathcal{O}(m)$ (m denotes the number of the degrees of freedom) when using Gaussian elimination to solve the matrix system with left-hand side matrix, $M + \eta K$. Adopting our splitting techniques, determining $\tilde{G}^{-1}$ needs $\mathcal{O}(m_x m_y)$ and $\mathcal{O}(m_x m_y m_z)$

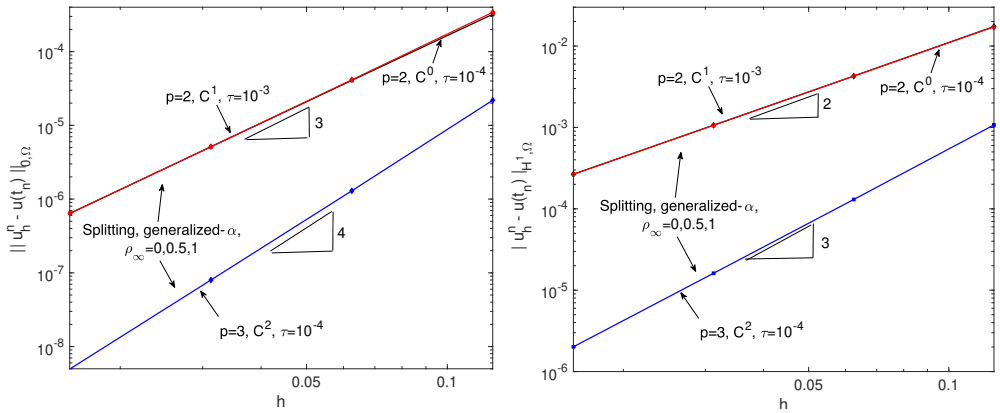


Figure 4.13: Spatial convergence of solution for split generalized-α method for $\rho_\infty = 0, 0.5, 1$ in L^2 norm & H^1 semi-norm for $p = 2, C^1$, $p = 3, C^2$ & $p = 2, C^0$ elements.

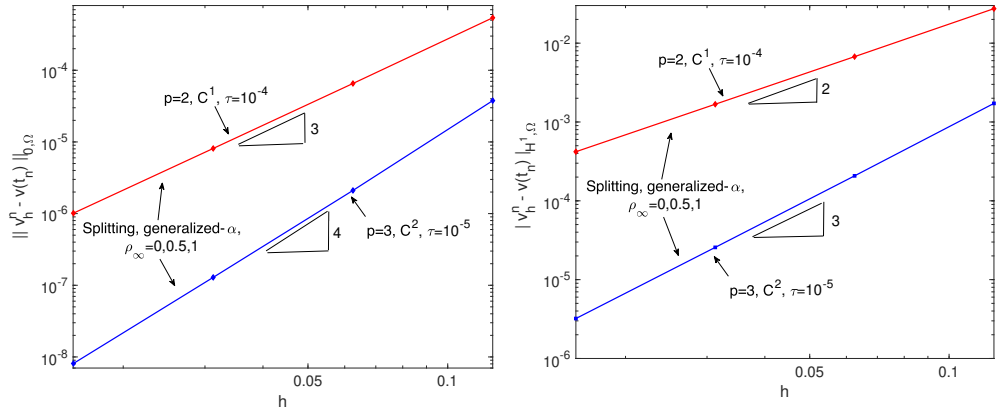


Figure 4.14: Spatial convergence of solution velocity for split generalized-α method for $\rho_\infty = 0, 0.5, 1$ in L^2 norm & H^1 semi-norm for $p = 2, C^1$, $p = 3, C^2$ & $p = 2, C^0$ elements.

for the 2D and 3D problems, respectively. Figure 4.12 shows the computational cost for the wave propagation problem (3.1.28). In this example, we utilize a direct solver (i.e., Gaussian elimination) for three space polynomials (i.e., C^0 & C^1 quadratic and C^2 cubic B-splines in each coordinate direction). This figure indicates the efficiency of the splitting scheme for solving the resulting matrix problems with linear cost.

Figure 4.13 presents the spatial convergence of the estimated displacement in the L^2 norm and H^1 semi-norm at the final time of 0.1 for fixed time-step sizes $\tau = 10^{-3}$ and $\tau = 10^{-4}$ exploiting C^1 quadratic and C^2 cubic elements, respectively. Additionally, we verify that the splitting technique for classical finite elements also provides the optimal convergence rates, $p + 1$ in L^2 norm, and p in H^1 norm. In this test, the splitting

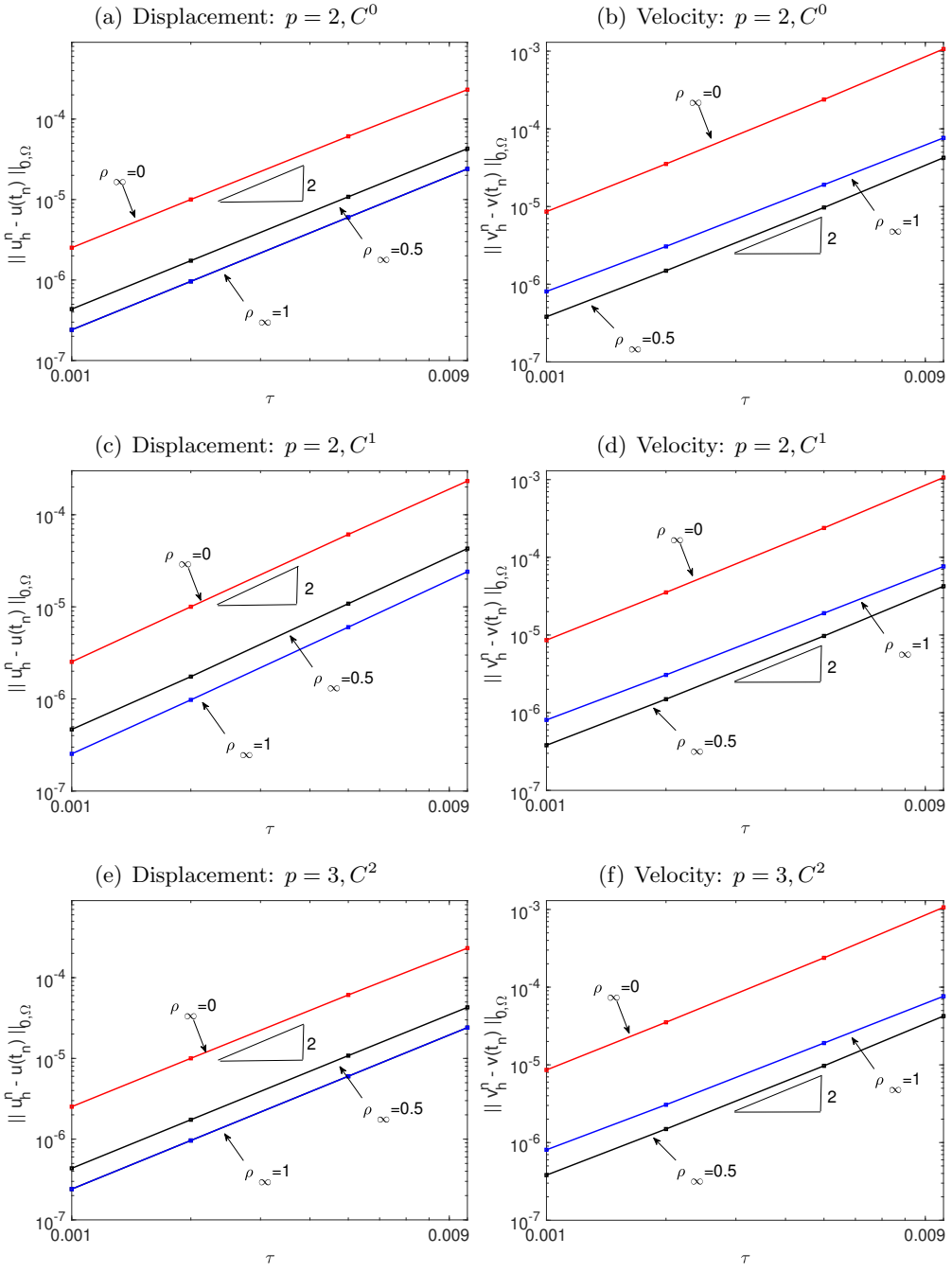


Figure 4.15: Temporal convergence of splitting method, in L^2 norm for $p = 2, C^0, n = 100$, $p = 2, C^1, n = 100$ & $p = 3, C^2, n = 64$ elements.

technique delivers the same error as the direct solution of the generalized-α method for $\rho_\infty = 0, 0.5, 1$. We show the velocity's spatial convergence rates for C^1 quadratic and C^2 cubic B-splines in figure 4.14.

In order to show the second-order accuracy in the temporal domain, we use $N = 100 \times 100$ quadratic elements for finite element and isogeometric analysis methods with and $N = 64 \times 64$ cubic elements for isogeometric. We set the final time $T = 0.1$. Figure 4.15 presents the results.

	Generalized-α method		Splitting	
τ	$\|u_h^n - u(t_n)\|_{0,\Omega}$	$\|v_h^n - v(t_n)\|_{0,\Omega}$	$\|u_h^n - u(t_n)\|_{0,\Omega}$	$\|v_h^n - v(t_n)\|_{0,\Omega}$
0.05	1.93^{-3}	1.41^{-2}	1.01^{-3}	1.45^{-2}
0.025	3.66^{-4}	5.07^{-3}	4.18^{-4}	4.80^{-3}
0.01	1.39^{-4}	7.91^{-4}	1.10^{-4}	7.53^{-4}
0.005	3.87^{-5}	2.07^{-4}	2.74^{-5}	1.73^{-4}
0.0025	9.73^{-6}	5.05^{-5}	6.85^{-6}	4.12^{-5}

Table 4.1: $L^2(\Omega)$ displacement & velocity errors for different time step sizes at final-time $T = 1$ with $\rho_\infty = 0$ and quadratic C^1 elements (time-dependent forcing, section 4.4.2.2).

4.4.2.2 Time-dependent forcing test

In this example, we approximate (3.1.28) with a time-dependent forcing function. We discretise the 3D domain $\Omega = (0,\, 1)^3 \in \mathbb{R}^3$ using 100 elements in each direction ($n_x = n_y = n_z = 100$) and account for the homogenous Dirichlet boundary conditions and initial conditions $u(.,\, 0) = v(.,\, 0) = 0$ with the forcing term and exact solution given by

$$
\begin{aligned}
f(x, y, z, t) &= \frac{\pi^2}{2}\left(3m^2 - (3m^2 - 4n^2)\cos(2n\pi t)\right)\sin(m\pi x)\sin(m\pi y)\sin(m\pi z), \\
u(x, y, z, t) &= \sin(m\pi x)\sin(m\pi y)\sin(m\pi z)\sin^2(n\pi t).
\end{aligned}
\tag{4.4.27}
$$

Setting $\rho_\infty = 0$, $m = 5$, $n = 2$ and deploying quadratic C^1 elements, we compare the results obtained using the generalized-α method (4.4.2) and the splitting method detailed in Algorithm 2. Table 4.1 reports the $L^2(\Omega)$ error of the displacement and velocity (i.e., $\|u_h^n - u(t_n)\|_{0,\Omega}$ and $\|v_h^n - v(t_n)\|_{0,\Omega}$) at the final-time $T = 1$.

4.4.2.3 Wave-packet problem

For the sake of completeness, we show the performance of our method for solving a wave-packet problem. A wave packet consists of an infinite set of component sinusoidal waves with different wavenumbers. They interfere constructively over a region of the spatial domain and destructively elsewhere. This localized behaviour of travelling waves has vast applications such as in seismic techniques, electrodynamics, and uncertainty analysis (see, e.g., [95]). We present the solution of a problem defined on the domain $\Omega = (-0.5,\, 1.5)^2 \in \mathbb{R}^2$ with the following given data:

$$\begin{aligned} u(x,y,0) &= \exp\left[15\left(x(i-15x)+y(i-15y)\right)\right], \\ u(x,y,0) &= 30(-i+15(x+y))\exp\left[15\left(x(i-15x)+y(i-15y)\right)\right], \\ f(x,y,t) &= 450(30t-30x+i)(30t-30y+i) \\ &\qquad \exp\left[-15\left(15(t-x)^2+15(t-y)^2+2it-ix-iy\right)\right], \end{aligned} \tag{4.4.28}$$

where i is the imaginary unit. We use quadratic C^1 elements with $n_x = n_y = 300$ and time-step $\tau = 10^{-3}$. Figures 4.16 and 4.17, respectively present the real and imaginary parts of the solution $u_h(\cdot,\, T)$ at $T = 0.05,\, 1.35$. Both the generalize-α method and our splitting technique deliver similar approximations.

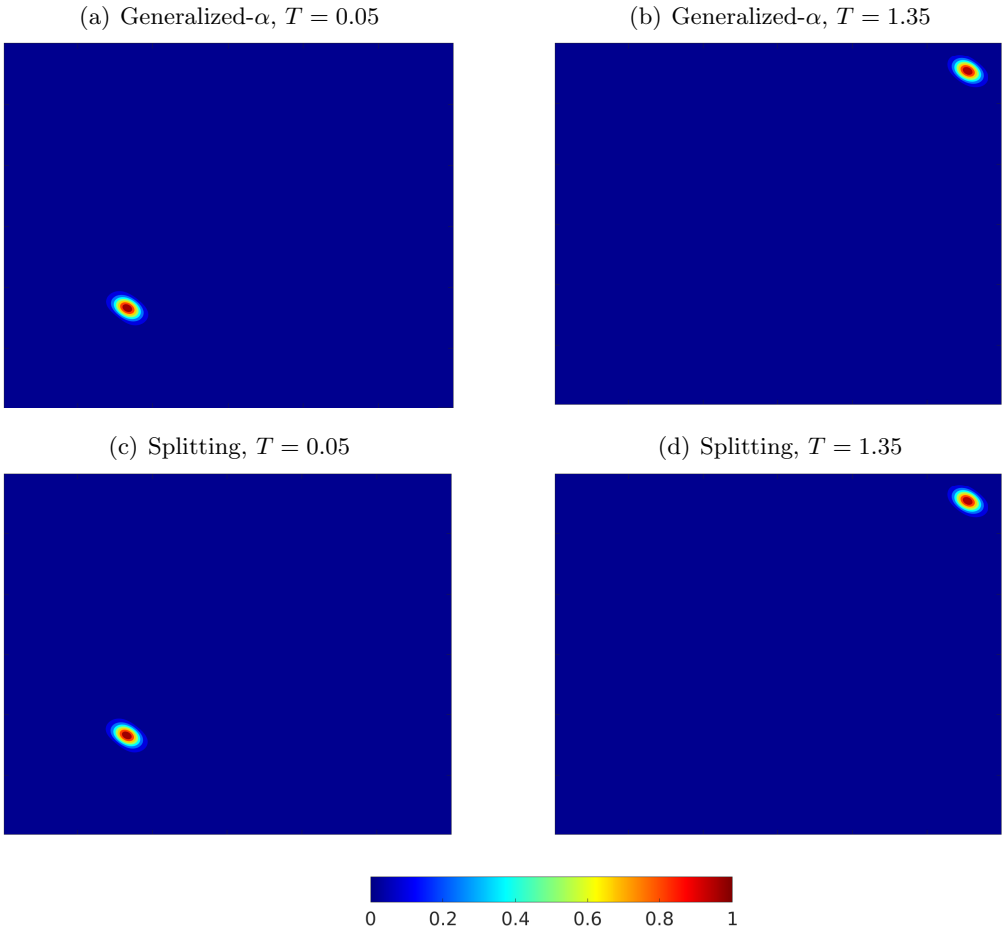


Figure 4.16: Comparison between the real part of the solutions obtained using the generalized-α method & our splitting technique in wave-packet problem.

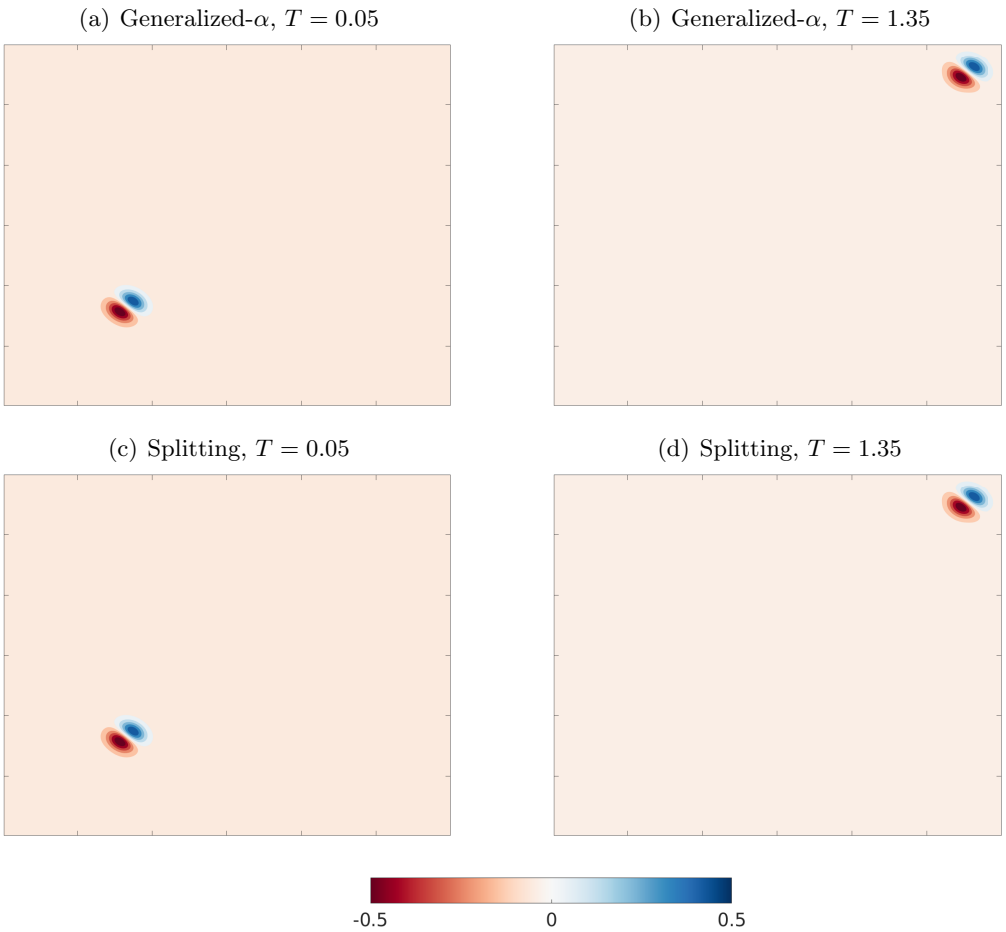


Figure 4.17: Approximate imaginary part in wave-packet problem.

5

Concluding remarks

The main results of the thesis can be summarized as follows:

- In this thesis, we investigate the localization phenomena in the folding of geological settings and we emphasize how nonlinear material properties and time-dependent responses (e.g., here, viscosity of the folded layer) can lead to spatial localization. In our models, the nonlinear material behaviour that is of interest is a nonlinear elastic softening-stiffening of a Winkler type foundation. We derive models based on the Swift-Hohenberg by minimizing the energy of the system. Next, we also include the viscous behaviour of the plate and propose a model accordingly. Lastly, to include transverse shear strains that gain importance by increasing the thickness of the plate, we develop a model employing higher-order shear train theory.

  We provide a detailed analysis of the dimension of the models and introduce the dimensionless versions. Later, we deliver a bound on the corresponding parameter to guarantee linear stability.

  Finally, we simulate the models using a robust finite element framework. For this, we use the method of lines and discretise the spatial domain using isogeometric analysis with sufficient regularity and the generalized-α method to discretise the temporal domain.

- We propose and analyse a new class of higher-order generalized-α methods with higher-order of accuracy in the temporal domain. In [10], we derive the model for solving hyperbolic problems with the second derivative with respect to time.

These problems have vast application in science and engineering such as wave propagation in rocks and elastic deformations. Then, we introduce our technique for estimating parabolic problems in [8, 10]. Our methods maintain all the attractive features of the original (second-order) generalized-α method. Similarly, our methods are a one-parameter family with dissipation control using a user-specified parameter ρ^{∞}.

Besides, our time-marching schemes are unconditionally stable for any high order of accuracy. Additionally, Our method for parabolic problems is A-stable for arbitrarily high accuracy. Furthermore, by setting $\rho^{\infty} = 0$, our method shows L-stability behaviour.

Lastly, we propose an explicit version of our high-order methods for solving hyperbolic problems. This allows us to approximate the models with less computational cost and able to deploy elements with high-order polynomial degrees. We also derived the corresponding CFL conditions and show that it is independent of the accuracy of the method (see, [12]).

- In the last part of this research, we develop a technique to decrease the computational cost of solving a matrix problem arising from spatial-temporal discretization. For this, we introduce a variational splitting that provides a linear cost solver with respect to the number of degrees of freedom for multi-dimensional parabolic problems [7]. Then, we extend our approach for approximating hyperbolic problems in [11]. This technique is very helpful in estimating problems that required high spatial resolution. For instance, we solve a wave-packet problem with very fine spatial elements. A wave packet is a short burst or envelope of localized wave action that travels as a unit and has a great interest to geologists.

  Next, we continue our work and derive and demonstrate the performance of our variational splitting scheme to simulate P-wave propagation problems as well as the linear elasticity problem [69]. That is, we decouple the vector problem using alternating triangular methods and build the spatial discretization on tensor-product spaces. We then use the Kronecker-product structure of the algebraic system to invert a sequence of implicit time steps with a cost proportional to the total number of degrees of freedom in the system. In this work, we particularly use high-order, smooth isogeometric basis functions in space.

In these papers, we prove theoretically and experimentally the unconditional stability of the methods. Also, we verify the second-order accuracy of the time schemes analytically and in our numerical experiments.

6

Appendix

.1 The coefficients of Cayley–Hamilton Theorem

For a given invertible matrix G, we can determine the coefficients $c_0 = (-1)^n \det(G)$ and other coefficients c_i, $i \in \{1, \cdots, n-1\}$ in terms complete exponential Bell polynomials B_l as [53]:

$$c_{n-l} = \frac{(-1)^l}{l!} B_l \left(s_1, -s_2, 2!s_3, \cdots, (-1)^{l-1}(l-1)!s_l \right), \tag{.1.1}$$

where s_l is the power sum of symmetric polynomials of the eigenvalues:

$$s_l = \sum_{i=1}^{l} \lambda_i^l = tr\left(G^l\right), \tag{.1.2}$$

with $tr\left(G^l\right)$ the trace of G^l. The l^{th} complete exponential Bell polynomial reads:

$$B_l(x_1, \cdots, x_l) = \sum_{m=1}^{l} B_{l,m}(x_1, \cdots, x_{l-m+1}), \tag{.1.3}$$

and defining the partial exponential Bell polynomials $B_{l,m}$ as:

$$B_{l,m}(x_1, \cdots, x_{l-m+1}) = \sum \frac{l!}{j_1! j_2! \cdots j_{l-m+1}!} (x_1)^{j_1} \left(\frac{x_1}{2}\right)^{j_2} \left(\frac{x_{l-m+1}}{(l-m+1)!}\right)^{j_{(l-m+1)}} \tag{.1.4}$$

where the sum is taken over all sequences $j_1, j_2, j_3, \cdots, j_{l-m+1}$ of non-negative integers such that these two conditions are satisfied:

$$\begin{aligned} j_1 + j_2 + \cdots + j_{l-m+1} &= m, \\ j_1 + 2j_2 + \cdots + (l-m+1)j_{l-m+1} &= l. \end{aligned} \tag{.1.5}$$

Following [55], it is also possible to determine B_l using the determinant as:

$$B_l(x_1,\cdots,x_l)=\det\begin{bmatrix} x_1 & x_2 & \frac{x_3}{2!} & \cdots & \frac{x_l}{(l-1)!} \\ -1 & x_1 & x_2 & \cdots & \frac{x_{l-1}}{(l-2)!} \\ 0 & -2 & x_1 & \cdots & \frac{x_{l-2}}{(l-3)!} \\ 0 & 0 & -3 & \cdots & \frac{x_{l-3}}{(l-4)!} \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & -(l-1) & x_1 \end{bmatrix}. \tag{.1.6}$$

For example, one can readily obtain

$$B_2(x_1,x_2)=x_1^2+x_2, \tag{.1.7}$$

$$B_3(x_1,x_2,x_3)=x_1^3+3x_1x_2+x_3, \tag{.1.8}$$

$$B_4(x_1,x_2,x_3,x_4)=x_1^4+6x_1^2x_2+4x_1x_3+3x_2^2+x_4, \tag{.1.9}$$

$$B_5(x_1,x_2,x_3,x_4,x_5)=x_1^5+10x_1^3x_2+15x_2^2x_1+10x_1^2x_3 \tag{.1.10}$$

$$+10x_3x_2+5x_4x_1+x_5, \tag{.1.11}$$

$$B_6(x_1,x_2,x_3,x_4,x_5,x_6)=x_1^6+15x_1^4x_2+20x_1^3x_3+45x_1^2x_2^2 \tag{.1.12}$$

$$+15x_2^3+60x_3x_2x_1+15x_1^2x_4+10x_3^2+15x_4x_2+6x_5x_1+x_6. \tag{.1.13}$$

Then, we using (.1.2) in (.1.7) and multiplying by U^{n-1}, we obtain (3.1.11). Similarly, introducing (.1.2) into (.1.9) and multiplying by U^{n-3} leads to (3.3.16).

.2 Linear computational cost solver

We assume matrix $\mathcal{M}=\mathcal{M}^x\otimes\mathcal{M}^y$ has a Kronecker-product structure. That is, each of the matrices $\mathcal{M}^\xi$ corresponds to the one-dimensional mass matrix in the direction ξ. Therefore, we can factorize the problem with linear cost with respect to the total number of degrees of freedom in the system.

These one-dimensional matrices have entries that correspond to the integrals of the multiplication of the one-dimensional B-spline basis functions. These B-spline basis functions have local support over $p+1$ elements, so the one-dimensional matrices $\mathcal{M}^x$, $\mathcal{M}^y$ have a banded structure.

$$\mathcal{M}^x_{ij}=0 \iff |i-j|>p \tag{.2.1}$$

$$\begin{bmatrix} \mathcal{M}^x_{11} & \mathcal{M}^x_{12} & \mathcal{M}^x_{13} & \mathcal{M}^x_{14} & 0 & 0 & \cdots & 0 \\ \mathcal{M}^x_{21} & \mathcal{M}^x_{22} & \mathcal{M}^x_{23} & \mathcal{M}^x_{24} & \mathcal{M}^x_{25} & 0 & \cdots & 0 \\ \mathcal{M}^x_{31} & \mathcal{M}^x_{32} & \mathcal{M}^x_{33} & \mathcal{M}^x_{34} & \mathcal{M}^x_{35} & \mathcal{M}^x_{36} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \dots & \mathcal{M}^x_{n(n-3)} & \mathcal{M}^x_{n(n-2)} & \mathcal{M}^x_{n(n-1)} & \mathcal{M}^x_{nn} \end{bmatrix}$$

where $\mathcal{M}^x_{ij} = (B^x_i, B^x_j)_{L^2}$. Same applies for $\mathcal{M}^y_{ij}$.

The Kronecker product structure of the matrix allows us to perform the following trick. Rather than solving a 3D problem, we can solve three one-dimensional problems with multiple right-hand-sides.

$$\begin{bmatrix} \mathcal{M}^x_{11} & \mathcal{M}^x_{12} & \mathcal{M}^x_{13} & \mathcal{M}^x_{14} & 0 & \cdots & 0 \\ \mathcal{M}^x_{21} & \mathcal{M}^x_{22} & \mathcal{M}^x_{23} & \mathcal{M}^x_{24} & \mathcal{M}^x_{25} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \mathcal{M}^x_{n(n-3)} & \mathcal{M}^x_{n(n-2)} & \mathcal{M}^x_{n(n-1)} & \mathcal{M}^x_{nn} \end{bmatrix}$$

$$\begin{bmatrix} x_{111} & x_{121} & \cdots & x_{1lm} \\ x_{211} & x_{221} & \cdots & x_{2lm} \\ \vdots & \vdots & \ddots & \vdots \\ x_{k11} & x_{k21} & \cdots & x_{klm} \end{bmatrix} = \begin{bmatrix} b_{111} & b_{121} & \cdots & b_{1lm} \\ b_{211} & b_{221} & \cdots & b_{2lm} \\ \vdots & \vdots & \ddots & \vdots \\ b_{k11} & b_{k21} & \cdots & b_{klm} \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{M}^y_{11} & \mathcal{M}^y_{12} & \mathcal{M}^y_{13} & \mathcal{M}^y_{14} & 0 & \cdots & 0 \\ \mathcal{M}^y_{21} & \mathcal{M}^y_{22} & \mathcal{M}^y_{23} & \mathcal{M}^y_{24} & \mathcal{M}^y_{25} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \mathcal{M}^y_{n(n-3)} & \mathcal{M}^y_{n(n-2)} & \mathcal{M}^y_{n(n-1)} & \mathcal{M}^y_{nn} \end{bmatrix}$$

$$\begin{bmatrix} y_{111} & y_{211} & \cdots & y_{k1m} \\ y_{121} & y_{211} & \cdots & y_{k2m} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1l1} & y_{1l1} & \cdots & y_{klm} \end{bmatrix} = \begin{bmatrix} x_{111} & x_{111} & \cdots & x_{k1m} \\ x_{121} & x_{211} & \cdots & x_{k2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1l1} & x_{2l1} & \cdots & x_{klm} \end{bmatrix}$$

$$\begin{bmatrix} \mathcal{M}^z_{11} & \mathcal{M}^z_{12} & \mathcal{M}^z_{13} & \mathcal{M}^z_{14} & 0 & \cdots & 0 \\ \mathcal{M}^z_{21} & \mathcal{M}^z_{22} & \mathcal{M}^z_{23} & \mathcal{M}^z_{24} & \mathcal{M}^z_{25} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & \mathcal{M}^z_{n(n-3)} & \mathcal{M}^z_{n(n-2)} & \mathcal{M}^z_{n(n-1)} & \mathcal{M}^z_{nn} \end{bmatrix}$$

$$\begin{bmatrix} z_{111} & z_{121} & \cdots & z_{1l1} \\ z_{212} & z_{222} & \cdots & z_{2l2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{k1m} & z_{k2m} & \cdots & z_{klm} \end{bmatrix} = \begin{bmatrix} y_{111} & y_{121} & \cdots & y_{1l1} \\ y_{212} & y_{222} & \cdots & y_{kl2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{k1m} & y_{k2m} & \cdots & y_{klm} \end{bmatrix}$$

where $\mathcal{M}^x_{ij} = (B^x_i, B^x_j)_{L^2}$ and $\mathcal{M}^y_{ij} = (B^y_i, B^y_j)_{L^2}$ and $\mathcal{M}^z_{ij} = (B^z_i, B^z_j)_{L^2}$. The first problem's dimensions are $n \times n$, where n is the number of B-spline basis functions along x axis, and we have ml right-hand-sides, where m is the number of B-spline basis functions along y axis, and l is the number of B-spline basis functions along z axis. The computational complexity of factorization of such a system is $O(n*m*l) = O(N)$ [71]. We have the analogous situation in the second problem, namely $m \times m$ system with $n*l$ right-hand-sides, which results in $O(m*n*l) = O(N)$ linear computational complexity, and in the third system we have $l \times l$ system with $n*m$ right-hand-sides, which results in $O(l*n*m) = O(N)$ linear computational complexity.

This strategy delivers a solution to the isogeometric L2 projection problem with linear $O(N)$ computational cost. This solution method improves on the standard direct solver cost estimates for and $O(N^2)$ in three-dimensions, see [100]) for the factorization of the global problem.

References

[1] M. Arnold and O. Brüls. Convergence of the generalized-α scheme for constrained mechanical systems. *Multibody System Dynamics*, **18**(2):185–202, 2007. 4

[2] U. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit runge-kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, **25**(2-3):151–167, 1997. 4

[3] U. M. Ascher and L. R. Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations*, **61**. Siam, 1998. 96, 99

[4] Y. Bazilevs, V. M. Calo, J. A. Cottrell, T. J. R. Hughes, A. Reali, and G. Scovazzi. Variational multiscale residual-based turbulence modeling for large eddy simulation of incompressible flows. *Computer Methods in Applied Mechanics and Engineering*, **197**(1-4):173–201, 2007. 4

[5] Y. Bazilevs, V. M. Calo, Y. Zhang, and T. J. R. Hughes. Isogeometric fluid–structure interaction analysis with applications to arterial blood flow. *Computational Mechanics*, **38**(4-5):310–322, 2006. 4

[6] L. E. Beckingham, C. Steefel, A. M. Swift, M. Voltolini, L. Yang, L. M. Anovitz, J. M. Sheets, D. R. Cole, T. J. Kneafsey, E. H. Mitnick, et al. Evaluation of accessible mineral surface areas for improved prediction of mineral reaction rates in porous media. *Geochimica et Cosmochimica Acta*, **205**:31–49, 2017. 1

[7] P. Behnoudfar, V. M. Calo, Q. Deng, and P. D. Minev. A variationally separable splitting for the generalized-α method for parabolic equations. *Inter-*

national Journal for Numerical Methods in Engineering, **121**(5):828–841, 2020. 40, 53, 62, 78, 84, 85, 105, 116

[8] P. Behnoudfar, Q. Deng, and V. M. Calo. High-order generalized-alpha method. *Applications in Engineering Science*, **4**:100021, 2020. 36, 37, 50, 58, 66, 116

[9] P. Behnoudfar, Q. Deng, and V. M. Calo. Higher-order generalized-α methods for parabolic problems. *arXiv preprint arXiv:2102.05910*, 2021. viii, 36, 60, 61, 62, 63, 66, 68, 70, 84, 85

[10] P. Behnoudfar, Q. Deng, and V. M. Calo. Higher-order generalized-α methods for hyperbolic problems. *Computer Methods in Applied Mechanics and Engineering*, **378**:113725, 2021. vii, viii, 36, 39, 40, 42, 43, 45, 47, 49, 50, 51, 58, 60, 99, 115, 116

[11] P. Behnoudfar, Q. Deng, and V. M. Calo. Split generalized-α method: A linear-cost solver for multi-dimensional second-order hyperbolic systems. *Computer Methods in Applied Mechanics and Engineering*, **376**:113656, 2021. 4, 78, 103, 104, 116

[12] Pouria Behnoudfar, Gabriele Loli, Alessandro Reali, Giancarlo Sangalli, and Victor M Calo. Explicit high-order generalized-α methods for isogeometric analysis of structural dynamics. *Computer Methods in Applied Mechanics and Engineering*, **389**:114344, 2022. viii, 36, 38, 45, 50, 51, 53, 54, 58, 116

[13] T. Belytschko, W. Liu, B. Moran, and K. Elkhodary. *Nonlinear finite elements for continua and structures*. John wiley & sons, 2013. 3

[14] M. A. Biot. Theory of folding of stratified viscoelastic media and its implications in tectonics and orogenesis. *Geological Society of America Bulletin*, **72**(11):1595–1620, 1961. 2, 30

[15] M. Bischoff, K. Bletzinger, W. Wall, and E. Ramm. Models and finite elements for thin-walled structures. *Encyclopedia of computational mechanics*, 2004. 7

[16] A. Bonelli and O. S. Bursi. Generalized-α methods for seismic structural testing. *Earthquake engineering & structural dynamics*, **33**(10):1067–1102, 2004. 4

[17] M. J. Borden, T. J. R. Hughes, C. M. Landis, and C. V. Verhoosel. A higher-order phase-field model for brittle fracture: Formulation and analysis within the isogeometric analysis framework. *Computer Methods in Applied Mechanics and Engineering*, **273**:100–118, 2014. 26

[18] O. Bruels and J. Golinval. The generalized-α method in mechatronic applications. *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik: Applied Mathematics and Mechanics*, **86**(10):748–758, 2006. 4

[19] O. Brüls, V. Acary, and A. Cardona. On the constraints formulation in the nonsmooth generalized-α method. In *Advanced Topics in Nonsmooth Dynamics*, pages 335–374. Springer, 2018. 4

[20] C. J. Budd and M. A. Peletier. Approximate self-similarity in models of geological folding. *SIAM Journal on Applied Mathematics*, **60**(3):990–1016, 2000. 2

[21] A. Buffa, C. De Falco, and G. Sangalli. Isogeometric analysis: new stable elements for the Stokes equation. *International Journal for Numerical Methods in Fluids*, 2010. 79

[22] A. Burbeau, P. Sagaut, and C. Bruneau. A problem-independent limiter for high-order runge–kutta discontinuous galerkin methods. *Journal of Computational Physics*, **169**(1):111–150, 2001. 4

[23] J. C. Butcher. *Numerical methods for ordinary differential equations*. John Wiley & Sons, 2016. 4, 92

[24] E. Cerda and L. Mahadevan. Geometry and physics of wrinkling. *Physical review letters*, **90**(7):074302, 2003. 12

[25] Y. Cheng, L. N. Wong, and C. Zou. Experimental study on the formation of faults from en-echelon fractures in carrara marble. *Engineering Geology*, **195**:312–326, 2015. 1

[26] J. Chung and G. M. Hulbert. A time integration algorithm for structural dynamics with improved numerical dissipation: the generalized-α method. *Journal of Applied Mechanics*, **60**(2):371–375, 1993. 3, 4, 36, 40, 46, 51, 62, 65, 104, 107

[27] S. P. Clavijo, L. Espath, A. Sarmiento, and V. M. Calo. A continuum theory for mineral solid solutions undergoing chemo-mechanical processes. *arXiv preprint arXiv:2101.04497*, 2021. 2

[28] J. A. Cottrell, T. J. R. Hughes, and Y. Bazilevs. *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons, 2009. 26, 45

[29] J. A. Cottrell, A. Reali, Y. Bazilevs, and T. J. R. Hughes. Isogeometric analysis of structural vibrations. *Computer Methods in Applied Mechanics and Engineering*, **195**(41):5257–5296, 2006. 26, 45

[30] W. D. Curtis, J. D. Logan, and W. A. Parker. Dimensional analysis and the pi theorem. *Linear Algebra and its Applications*, **47**:117–126, 1982. 16

[31] G. Dahlquist. A special stability problem for linear multistep methods. *BIT*, **3**:27–43, 1963. 102

[32] C. De Boor. *A practical guide to splines*, **27**. Springer-Verlag New York, 1978. 79

[33] J. Doetsch, V. S. Gischig, L. Villiger, H. Krietsch, M. Nejati, F. Amann, M. Jalali, C. Madonna, H. Maurer, S. Wiemer, et al. Subsurface fluid pressure and rock deformation monitoring using seismic velocity observations. *Geophysical Research Letters*, **45**(19):10–389, 2018. 1

[34] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, **82**(2):421–439, 1956. 5

[35] D. Durney. Solution-transfer, an important geological deformation mechanism. *Nature*, **235**(5337):315–317, 1972. 1

[36] B. Echebarria, R. Folch, A. Karma, and M. Plapp. Quantitative phase-field model of alloy solidification. *Physical review E*, **70**(6):061604, 2004. 2

[37] S. Erlicher, L. Bonaventura, and O. S. Bursi. The analysis of the generalized-α method for non-linear dynamic problems. *Computational mechanics*, **28**(2):83–104, 2002. 4

[38] L. Espath, A. F. Sarmiento, P. Vignal, B. Varga, A. Cortes, L. Dalcin, and V. M. Calo. Energy exchange analysis in droplet dynamics via the navier–stokes–cahn–hilliard model. *Journal of Fluid Mechanics*, **797**:389–430, 2016. 26

[39] J. A. Evans and T. J. R. Hughes. Isogeometric divergence-conforming b-splines for the Darcy–Stokes–Brinkman equations. *Mathematical Models and Methods in Applied Sciences*, **23**(04):671–741, 2013. 79

[40] K. F. Evans, H. Moriya, H. Niitsuma, R. H. Jones, W. S. Phillips, A. Genter, J. Sausse, R. Jung, and R. Baria. Microseismicity and permeability enhancement of hydrogeologic structures during massive fluid injections into granite at 3 km depth at the soultz hdr site. *Geophysical Journal International*, **160**(1):388–412, 2005. 1

[41] R. G. Faulkner. Non-equilibrium grain-boundary segregation in austenitic alloys. *Journal of Materials Science*, **16**(2):373–383, 1981. 73

[42] R. C. Fletcher. Wavelength selection in the folding of a single layer with power-law rheology. *American Journal of Science*, **274**(9):1029–1043, 1974. 2

[43] L. Gao and V. M. Calo. Fast isogeometric solvers for explicit dynamics. *Computer Methods in Applied Mechanics and Engineering*, **274**:19–41, 2014. 5, 77

[44] L. Gao and V. M. Calo. Preconditioners based on the alternating-direction-implicit algorithm for the 2d steady-state diffusion equation with orthotropic heterogeneous coefficients. *Journal of Computational and Applied Mathematics*, **273**:274–295, 2015. 5, 77

[45] I. Gelfand and R. A. Silverman. *Calculus of variations*. Courier Corporation, 2000. 9

[46] H. Gómez, V. M. Calo, Y. Bazilevs, and T. J. R. Hughes. Isogeometric analysis of the cahn–hilliard phase-field model. *Computer methods in applied mechanics and engineering*, **197**(49-50):4333–4352, 2008. 4, 26, 29

[47] H. Gomez, T. J. R. Hughes, X. Nogueira, and V. M. Calo. Isogeometric analysis of the isothermal navier–stokes–korteweg equations. *Computer Methods in Applied Mechanics and Engineering*, **199**(25-28):1828–1840, 2010. 4

[48] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, **14**. Springer, 2010. 47, 48, 53, 68, 69

[49] H. Hanche-Olsen. Buckingham's pi-theorem. *NTNU: http://www. math. ntnu. no/˜ hanche/notes/buckingham/buckingham-a4. pdf*, 2004. 16

[50] H. M. Hilber, T. J. R. Hughes, and R. L. Taylor. Improved numerical dissipation for time integration algorithms in structural dynamics. *Earthquake Engineering & Structural Dynamics*, **5**(3):283–292, 1977. 4

[51] B. Hobbs and A. Ord. Localized and chaotic folding: the role of axial plane structures. *Phil. Trans. R. Soc. A*, **370**(1965):1966–2009, 2012. 2

[52] B. Hobbs, K. Regenauer-Lieb, and A. Ord. Folding with thermal–mechanical feedback. *Journal of Structural Geology*, **30**(12):1572–1592, 2008. 2

[53] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 1990. 61, 65, 83, 119

[54] S. M. Houghton and E. Knobloch. Swift-hohenberg equation with broken cubic-quintic nonlinearity. *Physical Review E*, **84**(1):016204, 2011. 3

[55] A. S. Householder. *The theory of matrices in numerical analysis*. Courier Corporation, 2013. 65, 120

[56] T. J. R. Hughes. *The finite element method: linear static and dynamic finite element analysis*. Courier Corporation, 2012. 38, 44, 57, 65, 80

[57] G. M. Hulbert and J. Chung. Explicit time integration algorithms for structural dynamics with optimal numerical dissipation. *Computer Methods in Applied Mechanics and Engineering*, **137**(2):175 – 188, 1996. 53

[58] G. W. Hunt, H. M. Bolt, and J. M. Thompson. Structural localization phenomena and the dynamical phase-space analogy. *Proc. R. Soc. Lond. A*, **425**(1869):245–267, 1989. 2

[59] A. Iserles. *A first course in the numerical analysis of differential equations*. Number 44. Cambridge university press, 2009. 96

[60] K. E. Jansen, C. H. Whiting, and G. M. Hulbert. A generalized-α method for integrating the filtered Navier–Stokes equations with a stabilized finite element method. *Computer Methods in Applied Mechanics and Engineering*, **190**(3-4):305–319, 2000. 4, 28, 29, 40, 57, 58, 60, 62, 65, 84

[61] A. Joets and R. Ribotta. Localized, time-dependent state in the convection of a nematic liquid crystal. *Physical review letters*, **60**(21):2164, 1988. 2

[62] A. M. Johnson and R. C. Fletcher. *Folding of viscous layers: mechanical analysis and interpretation of structures in deformed rock*. Columbia University Press, 1994. 1

[63] A. Karma, D. A. Kessler, and H. Levine. Phase-field model of mode iii dynamic fracture. *Physical Review Letters*, **87**(4):045501, 2001. 2

[64] S. G. Kim, W. T. Kim, and T. Suzuki. Phase-field model for binary alloys. *Physical review e*, **60**(6):7186, 1999. 2

[65] P. Kolodner. Coexisting traveling waves and steady rolls in binary-fluid convection. *Physical Review E*, **48**(2):R665, 1993. 2

[66] C. Kuhn and R. Müller. A continuum phase field model for fracture. *Engineering Fracture Mechanics*, **77**(18):3625–3634, 2010. 2

[67] K. Lee, W. D. McCormick, J. E. Pearson, and H. L. Swinney. Experimental observation of self-replicating spots in a reaction–diffusion system. *Nature*, **369**(6477):215, 1994. 2

[68] P. Lejcek. *Grain boundary segregation in metals*, **136**. Springer Science & Business Media, 2010. 73

[69] M. Łoś, P. Behnoudfar, M. Paszyński, and V. M. Calo. Fast isogeometric solvers for hyperbolic wave propagation problems. *Computers & Mathematics with Applications*, **80**(1):109–120, 2020. 5, 78, 116

[70] M. Łoś, M. Paszyński, A. Kłusek, and W. Dzwinel. Application of fast isogeometric l2 projection solver for tumor growth simulations. *Computer Methods in Applied Mechanics and Engineering*, **316**:1257–1269, 2017. 5, 77, 81, 88, 103

[71] M. Łoś, M. Woźniak, M. Paszyński, L. Dalcin, and V. M. Calo. Dynamics with matrices possessing Kronecker product structure. *Procedia Computer Science*, **51**:286–295, 2015. 5, 77, 81, 103, 122

[72] A. E. Love. *A treatise on the mathematical theory of elasticity*. Cambridge university press, 2013. 12

[73] G. Marchuk. Splitting and alternating direction methods. *Handbook of numerical analysis*, **1**:197–462, 1990. 5

[74] F. Niordson. *Shell theory*. Elsevier, 2012. 3

[75] T. Ohta and K. Kawasaki. Equilibrium morphology of block copolymer melts. *Macromolecules*, **19**(10):2621–2632, 1986. 73

[76] J. M. Ortega. *Numerical analysis: a second course*. SIAM, 1990. 93

[77] D. W. Peaceman and H. H. Rachford. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for industrial and Applied Mathematics*, **3**(1):28–41, 1955. 5

[78] L. Piegl and W. Tiller. *The NURBS book*. Springer Science & Business Media, 1997. 79

[79] S. Pirkl, P. Ribiere, and P. Oswald. Forming process and stability of bubble domains in dielectrically positive cholesteric liquid crystals. *Liquid Crystals*, **13**(3):413–425, 1993. 2

[80] N. J. Price and J. W. Cosgrove. *Analysis of geological structures*. Cambridge University Press, 1990. 1

[81] J. N. Reddy. *Theory and analysis of elastic plates and shells*. CRC press, 2006. 3, 7

[82] E. H. Rutter. A discussion on natural strain and geological structure-the kinetics of rock deformation by pressure solution. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **283**(1312):203–219, 1976. 1

[83] Bruno S. An analysis of operator splitting techniques in the stiff case. *Journal of Computational Physics*, **161**(1):140 – 168, 2000. 5

[84] A. A. Samarskii, P. P. Matus, and P. N. Vabishchevich. *Difference schemes with operator factors*. Kluwer Academic Publisher, 2002. 5

[85] A. F. Sarmiento, L. Espath, P. Vignal, L. Dalcin, M. Parsani, and V. M. Calo. An energy-stable generalized-α method for the swift–hohenberg equation. *Journal of Computational and Applied Mathematics*, **344**:836–851, 2018. 4

[86] S. M. Schmalholz and Y.Y. Podladchikov. Finite amplitude folding: transition from exponential to layer length controlled growth. *Earth and Planetary Science Letters*, **179**(2):363–377, 2000. 2

[87] L. F. Shampine. Computer solution of ordinary differential equations. *The Initial Value Problem*, 1975. 77

[88] L. F. Shampine and C. W. Gear. A user's view of solving stiff ordinary differential equations. *SIAM review*, **21**(1):1–17, 1979. 77, 96

[89] J. Sherwin and W. M. Chapple. Wavelengths of single-layer folds; a comparison between theory and observation. *American Journal of Science*, **266**(3):167–179, 1968. 2

[90] R. B. Smith. Unified theory of the onset of folding, boudinage, and mullion structure. *Geological Society of America Bulletin*, **86**(11):1601–1609, 1975. 2

[91] R. B. Smith. Formation of folds, boudinage, and mullions in non-newtonian materials. *Geological Society of America Bulletin*, **88**(2):312–320, 1977. 2

[92] I. Steinbach. Phase-field model for microstructure evolution at the mesoscopic scale. *Annual Review of Materials Research*, **43**:89–107, 2013. 2

[93] E. Süli and D. F. Mayers. *An introduction to numerical analysis*. Cambridge university press, 2003. 55, 98, 99

[94] R. Szilard. *Theories and applications of plate analysis: classical, numerical and engineering methods*. John Wiley & Sons, 2004. 7

[95] J. Tate, T. Auguste, H. G. Muller, P. Salières, P. Agostini, and L. F. DiMauro. Scaling of wave-packet dynamics in an intense midinfrared field. *Physical Review Letters*, **98**(1):013901, 2007. 112

[96] V. Thomée. *Galerkin finite element methods for parabolic problems*, **1054**. Springer, 1984. 87

[97] S. P. Timoshenko and S. Woinowsky-Krieger. *Theory of plates and shells*. McGraw-hill, 1959. 7

[98] P. N. Vabishchevich. *Additive operator-difference schemes: Splitting schemes*. Walter de Gruyter, 2013. 5

[99] P. Vignal, N. Collier, L. Dalcin, D. L. Brown, and V. M. Calo. An energy-stable time-integrator for phase-field models. *Computer Methods in Applied Mechanics and Engineering*, **316**:1179–1214, 2017. 26

[100] P. Vignal, L. Dalcin, D. L. Brown, N. Collier, and V. M. Calo. An energy-stable convex splitting for the phase-field crystal equation. *Computers & Structures*, **158**:355–368, 2015. 29, 122

[101] N. Wawrzenitz, A. Krohe, D. Rhede, and R. Romer. Dating rock deformation with monazite: The impact of dissolution precipitation creep. *Lithos*, **134**:52–74, 2012. 1

[102] R. Weinstock. *Calculus of variations: with applications to physics and engineering*. Courier Corporation, 1974. 15, 21

[103] A. A. Wheeler, W. J. Boettinger, and G. B. McFadden. Phase-field model for isothermal phase transitions in binary alloys. *Physical Review A*, **45**(10):7424, 1992. 2

[104] L. Wong and H. H. Einstein. Systematic evaluation of cracking behavior in specimens containing single flaws under uniaxial compression. *International Journal of Rock Mechanics and Mining Sciences*, **46**(2):239–249, 2009. 1

[105] W. L. Wood, M. Bossak, and O. C. Zienkiewicz. An alpha modification of Newmark's method. *International Journal for Numerical Methods in Engineering*, **15**(10):1562–1566, 1980. 4

[106] H. Yasuhara and D. Elsworth. Compaction of a rock fracture moderated by competing roles of stress corrosion and pressure solution. *Pure and Applied Geophysics*, **165**(7):1289–1306, 2008. 1

[107] Y. Zhang, L. N. Wong, and K. Chan. An extended grain-based model accounting for microstructures in rock deformation. *Journal of Geophysical Research: Solid Earth*, **124**(1):125–148, 2019. 1